

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2002, 01:11:34 ; Search time 181.2 seconds  
(without alignments)  
194.250 Million cell updates/sec

Title: US-09-462-480-5

Perfect score: 492

Sequence: 1 MAEMKTDAAATLGAGNFR.....VOYSRADEEQOALSSOMGF 100

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 3502263 seqs, 351980561 residues

Total number of hits satisfying chosen parameters: 3502263

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending\_Patents\_AA\_Main:\*

1: /cgn2\_6/ptodata/2/paa/PCTUS\_COMB.pcp.\*  
2: /cgn2\_6/ptodata/2/paa/US06\_COMB.pcp.\*  
3: /cgn2\_6/ptodata/2/paa/US07\_COMB.pcp.\*  
4: /cgn2\_6/ptodata/2/paa/US080\_COMB.pcp.\*  
5: /cgn2\_6/ptodata/2/paa/US081\_COMB.pcp.\*  
6: /cgn2\_6/ptodata/2/paa/US082\_COMB.pcp.\*  
7: /cgn2\_6/ptodata/2/paa/US083\_COMB.pcp.\*  
8: /cgn2\_6/ptodata/2/paa/US084\_COMB.pcp.\*  
9: /cgn2\_6/ptodata/2/paa/US085\_COMB.pcp.\*  
10: /cgn2\_6/ptodata/2/paa/US086\_COMB.pcp.\*  
11: /cgn2\_6/ptodata/2/paa/US087\_COMB.pcp.\*  
12: /cgn2\_6/ptodata/2/paa/US088\_COMB.pcp.\*  
13: /cgn2\_6/ptodata/2/paa/US089\_COMB.pcp.\*  
14: /cgn2\_6/ptodata/2/paa/US090\_COMB.pcp.\*  
15: /cgn2\_6/ptodata/2/paa/US091\_COMB.pcp.\*  
16: /cgn2\_6/ptodata/2/paa/US092\_COMB.pcp.\*  
17: /cgn2\_6/ptodata/2/paa/US093\_COMB.pcp.\*  
18: /cgn2\_6/ptodata/2/paa/US094\_COMB.pcp.\*  
19: /cgn2\_6/ptodata/2/paa/US095\_COMB.pcp.\*  
20: /cgn2\_6/ptodata/2/paa/US096\_COMB.pcp.\*  
21: /cgn2\_6/ptodata/2/paa/US097\_COMB.pcp.\*  
22: /cgn2\_6/ptodata/2/paa/US098\_COMB.pcp.\*  
23: /cgn2\_6/ptodata/2/paa/US099\_COMB.pcp.\*  
24: /cgn2\_6/ptodata/2/paa/US100\_COMB.pcp.\*  
25: /cgn2\_6/ptodata/2/paa/US101\_COMB.pcp.\*  
26: /cgn2\_6/ptodata/2/paa/US60\_COMB.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	492	100.0	100	15	US-09-116-492-5
2	492	100.0	100	18	US-09-462-480-5
3	486	98.8	100	1	PCT-US99-03265-110
4	486	98.8	100	1	PCT-US99-03268-115
5	486	98.8	100	11	US-08-729-622-110
6	486	98.8	100	11	US-08-730-510-115
7	486	98.8	100	13	US-08-942-341-110

8	486	98.8	100	13	US-08-942-578-115
9	486	98.8	100	14	US-09-024-753-110
10	486	98.8	100	14	US-09-025-197-115
11	486	98.8	100	14	US-09-072-596-110
12	486	98.8	100	14	US-09-072-967-115
13	486	98.8	100	20	US-09-615-947-1
14	486	98.8	100	21	US-09-724-685-115
15	486	98.8	100	24	US-10-084-843-115
16	481	97.8	802	1	PCT-US99-03265-209
17	481	97.8	802	1	PCT-US99-03268-346
18	481	97.8	802	1	PCT-US99-03268-214
19	481	97.8	802	1	PCT-US99-03268-351
20	481	97.8	802	13	US-08-942-341-209
21	481	97.8	802	13	US-08-942-578-214
22	481	97.8	802	14	US-09-024-753-209
23	481	97.8	802	14	US-09-025-197-214
24	481	97.8	802	14	US-09-072-596-209
25	481	97.8	802	14	US-09-072-596-346
26	481	97.8	802	14	US-09-072-967-214
27	481	97.8	802	14	US-09-072-967-351
28	481	97.8	802	16	US-09-287-849-10
29	481	97.8	802	24	US-10-084-843-214
30	481	97.8	802	24	US-10-084-843-351
31	462	93.9	95	1	PCT-US99-03265-88
32	462	93.9	95	1	PCT-US99-03268-88
33	462	93.9	95	10	US-08-658-800-89
34	462	93.9	95	10	US-08-659-683-88
35	462	93.9	95	10	US-08-680-573-89
36	462	93.9	95	10	US-08-680-574-88
37	462	93.9	95	11	US-08-729-622-89
38	462	93.9	95	11	US-08-730-510-88
39	462	93.9	95	13	US-08-942-341-89
40	462	93.9	95	13	US-08-942-578-88
41	462	93.9	95	14	US-09-024-753-89
42	462	93.9	95	14	US-09-025-197-88
43	462	93.9	95	14	US-09-072-596-89
44	462	93.9	95	14	US-09-072-967-88
45	462	93.9	95	21	US-09-724-685-88

#### ALIGNMENTS

RESULT 1  
US-09-116-492-5  
; Sequence 5, Application US/09116492  
; GENERAL INFORMATION:  
; APPLICANT: GICQUEL, BRIGITTE  
; APPLICANT: BERTHET, FRANCOIS-XAVIER  
; APPLICANT: ANDERSEN, PETER  
; APPLICANT: RASMUSSEN, PETER B  
; TITLE OF INVENTION: POLYNUCLEOTIDE FUNCTIONALLY CODING FOR THE LHP PROTEIN  
; TITLE OF INVENTION: FROM MYCOBACTERIUM TUBERCULOSIS, ITS BIOLOGICALLY  
; TITLE OF INVENTION: ACTIVE DERIVATIVE FRAGMENTS, AS WELL AS METHODS USING  
; TITLE OF INVENTION: THE SAME  
; FILE REFERENCE: 0660-0137-27X  
; CURRENT APPLICATION NUMBER: US/09/116,492  
; CURRENT FILING DATE: 1998-07-16  
; EARLIER APPLICATION NUMBER: 60/052,631  
; EARLIER FILING DATE: 1997-07-16  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 100  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis  
US-09-116-492-5

Query Match 100.0%; Score 492; DB 15; Length 100;  
Best Local Similarity 100.0%; Pred. No. 3.8e-44;  
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAEMKTDATLGOEAGNFERISGLTKTQIDQVESTAGSLQGWGAGTAAGTAQAQAAVVRFE 60  
Db 1 MAEMKTDATLGOEAGNFERISGLTKTQIDQVESTAGSLQGWGAGTAAGTAQAQAAVVRFE 60  
QY 61 AANKQKQELDEISTNIRQAGVQYSRADEEQOQALSSQMGF 100  
Db 61 AANKQKQELDEISTNIRQAGVQYSRADEEQOQALSSQMGF 100

## RESULT

US-09-462-480-5  
; Sequence 5, Application US/09462480  
; GENERAL INFORMATION:  
; APPLICANT: GICQUEL, BRIGITTE  
; APPLICANT: BERTHET, FRANCIS-XAVIER  
; APPLICANT: ANDERSEN, PETER  
; APPLICANT: RASMUSSEN, PETER BIRK  
; TITLE OF INVENTION: POLYNUCLEOTIDE FUNCTIONALLY CODING FOR THE LHP PROTEIN FROM MYC  
; TITLE OF INVENTION: TUBERCULOSIS, ITS BIOLOGICALLY ACTIVE DERIVATIVE FRAGMENTS, AS W  
; FILE REFERENCE: 0660-0165-0XPCT  
; CURRENT APPLICATION NUMBER: US/09/462,480  
; PRIOR FILING DATE: 2000-03-06  
; PRIOR APPLICATION NUMBER: PCT/TB98/01091  
; PRIOR FILING DATE: 1998-07-16  
; PRIOR APPLICATION NUMBER: 60/052,631  
; PRIOR FILING DATE: 1997-07-16  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 5  
; LENGTH: 100  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis  
US-09-462-480-5

Query Match 100.0%; Score 492; DB 18; Length 100;  
Best Local Similarity 100.0%; Pred. No. 3.8e-44;  
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAEMKTDATLGOEAGNFERISGLTKTQIDQVESTAGSLQGWGAGTAAGTAQAQAAVVRFE 60  
Db 1 MAEMKTDATLGOEAGNFERISGLTKTQIDQVESTAGSLQGWGAGTAAGTAQAQAAVVRFE 60  
QY 61 AANKQKQELDEISTNIRQAGVQYSRADEEQOQALSSQMGF 100  
Db 61 AANKQKQELDEISTNIRQAGVQYSRADEEQOQALSSQMGF 100

## RESULT

PCT-US99-03265-110  
; Sequence 110, Application PC/TUS9903265  
; GENERAL INFORMATION:  
; APPLICANT: Corixa Corporation  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF  
; TITLE OF INVENTION: TUBERCULOSIS  
; NUMBER OF SEQUENCES: 350  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds, LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036-2811  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US99/03265  
; FILING DATE: 17-FEB-1999  
; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 09/024,753  
; FILING DATE: 18-FEB-1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Laura A. Coruzzi  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 9532-0023-228  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-493-4935  
; TELEFAX: 650-493-5556  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 110:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 100 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
PCT-US99-03265-110

Query Match 98.8%; Score 486; DB 1; Length 100;  
Best Local Similarity 99.0%; Pred. No. 1.7e-43;  
Matches 99; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MAEMKTDATLGOEAGNFERISGLTKTQIDQVESTAGSLQGWGAGTAAGTAQAQAAVVRFE 60  
Db 1 MAEMKTDATLGOEAGNFERISGLTKTQIDQVESTAGSLQGWGAGTAAGTAQAQAAVVRFE 60  
QY 61 AANKQKQELDEISTNIRQAGVQYSRADEEQOQALSSQMGF 100  
Db 61 AANKQKQELDEISTNIRQAGVQYSRADEEQOQALSSQMGF 100

## RESULT

PCT-US99-03268-115  
; Sequence 115, Application PC/TUS9903268  
; GENERAL INFORMATION:  
; APPLICANT: Corixa Corporation  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY  
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS  
; NUMBER OF SEQUENCES: 355  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds, LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036-2811  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US99/03268  
; FILING DATE: 17-FEB-1999  
; CLASSIFICATION: 406  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 09/025,197  
; FILING DATE: 18-FEB-1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura, A.  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 9532-0013-228  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-493-4935  
; TELEFAX: 650-493-5556  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 115:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 100 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:

TOPOLOGY: linear  
PCT-US99-03268-115

Query Match 98.8%; Score 486; DB 1; Length 100;  
Best Local Similarity 99.0%; Pred. No. 1.7e-43;  
Matches 99; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAEMKTDATLGOAGNFERISGDLKTQIDQVESTAGSLQGWGAGTAQAQAAVVRQOE 60  
|||||

Db 1 MAEMKTDATLGOAGNFERISGDLKTQIDQVESTAGSLQGWGAGTAQAQAAVVRQOE 60  
|||||

QY 61 AANKQKQELDEISTNIRQAGVQYSRADDEEQQALSSQMGF 100  
|||||

Db 61 AANKQKQELDEISTNIRQAGVQYSRADDEEQQALSSQMGF 100  
|||||

RESULT 5

US-08-729-622-110  
; Sequence 110, Application US/08729622  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Campos-Neto, Antonio  
; APPLICANT: Houghton, Raymond  
; APPLICANT: Vedvick, Thomas H.  
; APPLICANT: Twardzik, David R.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF  
; TUBERCULOSIS  
; NUMBER OF SEQUENCES: 132  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/729,622  
; FILING DATE: 11-OCT-1996  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.417C5  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 110:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 100 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
US-08-729-622-110

Query Match 98.8%; Score 486; DB 1; Length 100;  
Best Local Similarity 99.0%; Pred. No. 1.7e-43;  
Matches 99; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAEMKTDATLGOAGNFERISGDLKTQIDQVESTAGSLQGWGAGTAQAQAAVVRQOE 60  
|||||

Db 1 MAEMKTDATLGOAGNFERISGDLKTQIDQVESTAGSLQGWGAGTAQAQAAVVRQOE 60  
|||||

QY 61 AANKQKQELDEISTNIRQAGVQYSRADDEEQQALSSQMGF 100  
|||||

Db 61 AANKQKQELDEISTNIRQAGVQYSRADDEEQQALSSQMGF 100

RESULT 6

US-08-730-510-115  
; Sequence 115, Application US/08730510  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Campos-Neto, Antonio  
; APPLICANT: Houghton, Raymond  
; APPLICANT: Vedvick, Thomas S.  
; APPLICANT: Twardzik, Daniel R.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY  
; TUBERCULOSIS  
; NUMBER OF SEQUENCES: 137  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/730,510  
; FILING DATE: 27-AUG-1996  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.411C5  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 115:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 100 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
US-08-730-510-115

Query Match 98.8%; Score 486; DB 1; Length 100;  
Best Local Similarity 99.0%; Pred. No. 1.7e-43;  
Matches 99; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAEMKTDATLGOAGNFERISGDLKTQIDQVESTAGSLQGWGAGTAQAQAAVVRQOE 60  
|||||

Db 1 MAEMKTDATLGOAGNFERISGDLKTQIDQVESTAGSLQGWGAGTAQAQAAVVRQOE 60  
|||||

QY 61 AANKQKQELDEISTNIRQAGVQYSRADDEEQQALSSQMGF 100  
|||||

Db 61 AANKQKQELDEISTNIRQAGVQYSRADDEEQQALSSQMGF 100  
|||||

RESULT 7

US-08-942-341-110  
; Sequence 110, Application US/08942341  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Campos-Neto, Antonio  
; APPLICANT: Houghton, Raymond  
; APPLICANT: Vedvick, Thomas S.  
; APPLICANT: Twardzik, Daniel R.

APPLICANT: Lodes, Michael J.  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS  
NUMBER OF SEQUENCES: 209  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED AND BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/942,341  
FILING DATE: 01-OCT-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.417C7  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 110:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 100 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
US-08-942-341-110

Query Match 98.8%; Score 486; DB 13; Length 100;  
Best Local Similarity 99.0%; Pred. No. 1.7e-43;  
Matches 99; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MAEMKTDAAATLGOEAGNFERISGDLKTDQIDVESTAGSLQGWGAGTAQAQAAVVRFOE 60  
|||||  
Db 1 MAEMKTDAAATLGOEAGNFERISGDLKTDQIDVESTAGSLQGWGAGTAQAQAAVVRFOE 60  
QY 61 AANKQKQELDEISTNIRAGVQYSRADEEQQALSSOMGF 100  
|||||  
Db 61 AANKQKQELDEISTNIRAGVQYSRADEEQQALSSOMGF 100

RESULT 8  
US-08-942-578-115  
Sequence 115, Application US/08942578  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Skeiky, Yasir A.W.  
APPLICANT: Dillon, Davin C.  
APPLICANT: Campos-Neto, Antonio  
APPLICANT: Houghton, Raymond  
APPLICANT: Vedvick, Thomas S.  
APPLICANT: Twardzik, Daniel R.  
APPLICANT: Lodes, Michael J.  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY  
NUMBER OF SEQUENCES: 214  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED AND BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/942,578  
FILING DATE: 01-OCT-1997  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.411C7  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 115:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 100 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
US-08-942-578-115  
Query Match 98.8%; Score 486; DB 13; Length 100;  
Best Local Similarity 99.0%; Pred. No. 1.7e-43;  
Matches 99; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MAEMKTDAAATLGOEAGNFERISGDLKTDQIDVESTAGSLQGWGAGTAQAQAAVVRFOE 60  
|||||  
Db 1 MAEMKTDAAATLGOEAGNFERISGDLKTDQIDVESTAGSLQGWGAGTAQAQAAVVRFOE 60  
QY 61 AANKQKQELDEISTNIRAGVQYSRADEEQQALSSOMGF 100  
|||||  
Db 61 AANKQKQELDEISTNIRAGVQYSRADEEQQALSSOMGF 100  
RESULT 9  
US-09-024-753-110  
Sequence 110, Application US/09024753  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Skeiky, Yasir A.W.  
APPLICANT: Dillon, Davin C.  
APPLICANT: Campos-Neto, Antonio  
APPLICANT: Houghton, Raymond  
APPLICANT: Vedvick, Thomas S.  
APPLICANT: Twardzik, Daniel R.  
APPLICANT: Lodes, Michael J.  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS  
NUMBER OF SEQUENCES: 236  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED AND BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/024,753  
FILING DATE: 18-FEB-1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.417C8  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 110:



SEQUENCE CHARACTERISTICS:  
LENGTH: 100 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
US-09-024-753-110

Query Match 98.8%; Score 486; DB 14; Length 100;  
Best Local Similarity 99.0%; Pred. No. 1.7e-43;  
Matches 99; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAEMKTDAAATLQAEAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAAYVRFQE 60  
Db 1 MAEMKTDAAATLQAEAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAAYVRFQE 60  
QY 61 AANKQKQELDELSTNIRQAGVQYSRADEEQQALSSQMGE 100  
Db 61 AANKQKQELDELSTNIRQAGVQYSRADEEQQALSSQMGE 100

RESULT 10  
US-09-025-197-115  
; Sequence 115, Application US/09025197  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Campos-Neto, Antonio  
; APPLICANT: Houghton, Raymond  
; APPLICANT: Vedvick, Thomas S.  
; APPLICANT: Twardzik, Daniel R.  
; APPLICANT: Lodes, Michael J.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY  
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS  
; NUMBER OF SEQUENCES: 241  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED AND BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; FILING DATE: 18-FEB-1998  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.411C8  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 115:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 100 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
US-09-025-197-115

Query Match 98.8%; Score 486; DB 14; Length 100;  
Best Local Similarity 99.0%; Pred. No. 1.7e-43;  
Matches 99; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MAEMKTDAAATLQAEAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAAYVRFQE 60

Db 1 MAEMKTDAAATLQAEAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAAYVRFQE 60  
QY 61 AANKQKQELDELSTNIRQAGVQYSRADEEQQALSSQMGE 100  
Db 61 AANKQKQELDELSTNIRQAGVQYSRADEEQQALSSQMGE 100

RESULT 11  
US-09-072-596-110  
; Sequence 110, Application US/09072596  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Campos-Neto, Antonio  
; APPLICANT: Houghton, Raymond  
; APPLICANT: Vedvick, Thomas S.  
; APPLICANT: Twardzik, Daniel R.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Hendrickson, Ronald C.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF  
; NUMBER OF SEQUENCES: 350  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED AND BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; FILING DATE: 05-MAY-1998  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.417C9  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 110:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 100 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
US-09-072-596-110

Query Match 98.8%; Score 486; DB 14; Length 100;  
Best Local Similarity 99.0%; Pred. No. 1.7e-43;  
Matches 99; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAEMKTDAAATLQAEAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAAYVRFQE 60  
Db 1 MAEMKTDAAATLQAEAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAAYVRFQE 60  
QY 61 AANKQKQELDELSTNIRQAGVQYSRADEEQQALSSQMGE 100  
Db 61 AANKQKQELDELSTNIRQAGVQYSRADEEQQALSSQMGE 100

RESULT 12  
US-09-072-967-115  
; Sequence 115, Application US/09072967  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.

APPLICANT: Skeiky, Yasir A.W.  
APPLICANT: Dillon, Davin C.  
APPLICANT: Campos-Neto, Antonio  
APPLICANT: Houghton, Raymond  
APPLICANT: Vedzick, Thomas S.  
APPLICANT: Twardzik, Daniel R.  
APPLICANT: Lodes, Michael J.  
APPLICANT: Hendrickson, Ronald C.  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY  
TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS  
NUMBER OF SEQUENCES: 355  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED AND BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/072,967  
FILING DATE: 05-MAY-1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.411C9  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 115:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 100 amino acids  
TYPE: amino acid  
STRANDEDNESS: linear  
TOPOLOGY: linear  
US-09-072-967-115

Query Match 98.8%; Score 486; DB 14; Length 100;  
Best Local Similarity 99.0%; Pred. No. 1.7e-43;  
Matches 99; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MAEMKTDATLQAEAGNFRISGLDKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOE 60  
|||||  
DB 1 MAEMKTDATLQAEAGNFRISGLDKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOE 60  
|||||  
QY 61 AANKQKQELDEISTNIRQAGVQYSRADDEQQQALSSQMGE 100  
|||||  
DB 61 AANKQKQELDEISTNIRQAGVQYSRADDEQQQALSSQMGE 100  
|||||  
RESULT 13  
US-09-615-947-1  
Sequence 1, Application US/09615947  
GENERAL INFORMATION:  
APPLICANT: ANDERSEN, Peter et al.  
TITLE OF INVENTION: Tuberculosis vaccine and diagnostics  
TITLE OF INVENTION: based on the Mycobacterium tuberculosis esat-6 gene family  
FILE REFERENCE: 0459-0462P  
CURRENT APPLICATION NUMBER: US/09/615,947  
CURRENT FILING DATE: 2001-06-04  
NUMBER OF SEQ ID NOS: 31  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 1  
LENGTH: 100  
TYPE: PRT  
ORGANISM: M.Tuberculosis  
US-09-615-947-1

Query Match 98.8%; Score 486; DB 20; Length 100;  
Best Local Similarity 99.0%; Pred. No. 1.7e-43;  
Matches 99; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MAEMKTDATLQAEAGNFRISGLDKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOE 60  
|||||  
DB 1 MAEMKTDATLQAEAGNFRISGLDKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOE 60  
|||||  
QY 61 AANKQKQELDEISTNIRQAGVQYSRADDEQQQALSSQMGE 100  
|||||  
DB 61 AANKQKQELDEISTNIRQAGVQYSRADDEQQQALSSQMGE 100  
|||||  
RESULT 14  
US-09-724-685-115  
Sequence 115, Application US/09724685  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
Skeiky, Yasir  
Dillon, Davin C.  
Campos-Neto, Antonio  
TITLE OF INVENTION: Compounds and Methods for  
Immunotherapy and Diagnosis of Tuberculosis  
NUMBER OF SEQUENCES: 155  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/724,685  
FILING DATE: 28-Nov-2000  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/523,436  
FILING DATE: 01-SEP-1995  
APPLICATION NUMBER: US 08/533,634  
FILING DATE: 22-SEP-1995  
APPLICATION NUMBER: US 08/620,874  
FILING DATE: 22-MAR-1996  
APPLICATION NUMBER: US 08/659,683  
FILING DATE: 05-JUN-1996  
APPLICATION NUMBER: US 08/680,574  
FILING DATE: 12-JUL-1996  
APPLICATION NUMBER: WO PCT/US96/14674  
FILING DATE: 30-AUG-1996  
APPLICATION NUMBER: US 08/730,511  
FILING DATE: 11-OCT-1996  
APPLICATION NUMBER: US 08/818,112  
FILING DATE: 13-MAR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Bastian, Kevin L.  
REGISTRATION NUMBER: 34,774  
REFERENCE/DOCKET NUMBER: 014058-008561US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 115:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 100 amino acids  
TYPE: amino acid  
STRANDEDNESS: <Unknown>  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 115:

US-09-724-685-115

Query Match 98.8%; Score 486; DB 21; Length 100;  
Best Local Similarity 99.0%; Pred. No. 1.7e-43;  
Matches 99; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAEMKTDATLGOEAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAAYVRFQE 60  
Db 1 MAEMKTDATLGOEAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAAYVRFQE 60  
QY 61 AANKKQKQELDEISTNIRQAGVQYSRADEEQQALSSQMGF 100  
Db 61 AANKKQKQELDEISTNIRQAGVQYSRADEEQQALSSQMGF 100

RESULT 15

US-10-084-843-115  
; Sequence 115, Application US/10084843  
; GENERAL INFORMATION:

APPLICANT: Reed, Steven G.  
Skeiky, Yasir A.W.  
Dillon, Davin C.  
Campos-Neto, Antonio  
Houghton, Raymond  
Vedvick, Thomas S.  
Twardzik, Daniel R.  
Lodes, Michael J.  
Hendrickson, Ronald C.  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY  
AND DIAGNOSIS OF TUBERCULOSIS

NUMBER OF SEQUENCES: 355  
CORRESPONDENCE ADDRESS:  
ADDRESSER: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
City: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/084,843  
FILING DATE: 25-Feb-2002  
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/072,967  
FILING DATE: 05-MAY-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.

REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.411C9  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 115:

SEQUENCE CHARACTERISTICS:  
LENGTH: 100 amino acids  
TYPE: amino acid  
STRANDEDNESS: <Unknown>  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 115:  
US-10-084-843-115

Query Match 98.8%; Score 486; DB 24; Length 100;  
Best Local Similarity 99.0%; Pred. No. 1.7e-43;  
Matches 99; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAEMKTDATLGOEAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAAYVRFQE 60

Db 1 MAEMKTDATLGOEAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAAYVRFQE 60  
QY 61 AANKKQKQELDEISTNIRQAGVQYSRADEEQQALSSQMGF 100  
Db 61 AANKKQKQELDEISTNIRQAGVQYSRADEEQQALSSQMGF 100

Search completed: July 22, 2002, 01:16:48  
Job time: 314 sec



OM of: US-09-462-480-4 to: A\_Geneseq\_032802:\* out\_format : pfs  
Date: Jul 22, 2002 1:22 AM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:  
-MODEL=framet.n2p.model -DEV=xlh  
-O=/cn2.1/USPTO.spool/US09462480/runat\_18072002\_164418\_19413/app\_query.fasta\_1.2850  
-DB=A\_Geneseq\_032802\_QFMT=fastan\_SUFFIX=rag -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-GAPOP=4.500 -GAPEXT=0.050 -YGAPOP=10.000 -XGAPEXT=0.500  
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blsum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -NODE=LOCAL -OUTFMT=pfs  
-NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09462480\_@CNCI\_1\_57 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

Search information block:  
Query: US-09-462-480-4  
Query length: 302  
Database: A\_Geneseq\_032802:\*  
Database sequences: 747574  
Database length: 111073796  
Search time (sec): 134.850000

score_list:	Strd Orig	zScore	EScore	Len	Documentation
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAV03705 +	492.00	905.18	2.9e-42		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW81697 +	486.00	894.12	1.2e-41		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW81706 +	486.00	894.12	1.2e-41		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW84339 +	486.00	894.12	1.2e-41		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW39136 +	486.00	894.12	1.2e-41		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW38993 +	486.00	894.12	1.2e-41		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:AAW35218 +	486.00	894.12	1.2e-41		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:AAW19845 +	486.00	894.12	1.2e-41		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW81746 +	481.00	865.85	5.6e-41		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW84379 +	481.00	865.85	5.6e-41		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW32023 +	481.00	865.85	5.6e-41		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW32634 +	481.00	865.85	5.6e-41		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW39176 +	481.00	865.85	5.6e-41		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW39081 +	481.00	865.85	5.6e-41		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW39033 +	481.00	865.85	5.6e-41		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:AAW01901 +	481.00	863.98	5.8e-41		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:AAW32444 +	482.00	850.36	3.4e-39		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:AAW32376 +	462.00	850.36	3.4e-39		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW81747 +	462.00	850.36	3.4e-39		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW84321 +	462.00	850.36	3.4e-39		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW32097 +	462.00	850.36	3.4e-39		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW39118 +	462.00	850.36	3.4e-39		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW38981 +	462.00	850.36	3.4e-39		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW32454 +	392.00	722.93	5.1e-32		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:AAW32386 +	392.00	722.93	5.1e-32		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW81707 +	392.00	722.93	5.1e-32		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW84340 +	392.00	722.93	5.1e-32		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW39137 +	392.00	722.93	5.1e-32		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW38994 +	392.00	722.93	5.1e-32		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW03706 +	247.00	460.20	3.6e-17		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAV03707 +	205.00	384.21	7.1e-13		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAV03712 +	137.00	262.61	6.3e-06		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:AAW32460 +	122.00	234.97	0.0002		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW81698 +	122.00	234.97	0.0002		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW39128 +	122.00	234.97	0.0002		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:AAW32458 +	121.00	233.46	0.0003		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW81696 +	121.00	233.46	0.0003		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW39126 +	121.00	233.46	0.0003		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:AAW32457 +	118.00	227.93	0.0006		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:AAW32459 +	118.00	227.93	0.0006		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW81695 +	118.00	227.93	0.0006		

/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW81697 + 118.00 227.93 0.00  
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW39125 + 118.00 227.93 0.00  
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW39127 + 118.00 227.93 0.00  
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:AAW32455 + 114.00 220.22 0.00  
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW81693 + 114.00 220.22 0.00

seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAV03705

seq\_documentation\_block:  
ID AAY03705 standard; Protein; 100 AA.

AC AAY03705;

DT 07-JUN-1999 (first entry)

DE M. tuberculosis LHP polypeptide.

KW ESAT-6 protein; antigenic protein; LHP; Mycobacterium tuberculosis;  
KW immunogen; vaccine; bacteria; tuberculosis; regulatory; synergistic;  
KW immune response.

OS Mycobacterium tuberculosis.

PN WO9904005-A1.

PD 28-JAN-1999.

PF 16-JUL-1998; 98WO-IB01091.

PR 16-JUL-1997; 97US-0052631.

PA (INSP ) INST PASTEUR.

PI (STAT-) STATENS SERUM INST.

PI Andersen P, Berthet F, Gicquel B, Rasmussen PB;

DR WPI; 1999-132249/11.

XX N-PSDB; AAX29168, AAX29171.

XX New nucleic acid containing regulator and LHP gene of Mycobacterium  
PT tuberculosis - useful in vaccines, for diagnosis, and for expression  
of heterologous proteins

XX Claim 17; Page 64; 88pp; English.

CC The present invention is directed to a polynucleotide carrying the  
CC regulatory expression signals of the ESAT-6 protein as well as an open  
CC reading frame coding for an antigenic protein LHP from Mycobacterium  
CC tuberculosis. Host cells comprising the polynucleotide are used for the  
CC recombinant expression of the protein. The recombinant polypeptide can be  
CC used as immunogens and vaccines, to protect against bacteria of the  
CC M. tuberculosis complex in humans or animals (the vaccines may include  
CC other immunogenic proteins of the bacteria or their fragments,  
CC specifically ESAT-6); and (b) for diagnosing tuberculosis infection by  
CC detection of specific antibodies. The regulatory region present in the  
CC polynucleotide may be used to express almost any heterologous protein in  
CC mycobacteria, particularly as a fusion with polyhistidine. The two  
CC proteins encoded by the polynucleotide, LHP and ESAT-6, are expected to  
CC provide a synergistic increase in ability to induce a protective immune  
CC response. The present sequence represents the LHP polypeptide.

XX Sequence 100 AA;

alignment\_scores:  
Quality: 492.00 Length: 100  
Ratio: 4.920 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-462-480-4 x AAY03705 ..

Align seg 1/1 to: AAY03705 from: 1 to: 100



CC from a M. tuberculosis strain H37Rv genomic library using a probe  
CC derived from clone Tb38-1 (see AAY4384). The invention relates to  
CC compositions and methods for diagnosing tuberculosis. It provides  
CC polypeptides (see AAW64291-W64379) comprising an antigenic portion of  
CC a soluble M. tuberculosis antigen, or an immunogenic portion of an  
CC M. tuberculosis antigen, as well as DNA sequences encoding such  
CC polypeptides, recombinant expression vectors and transformed or  
CC transfected host cells. Also claimed are methods and diagnostic  
CC kits for detecting M. tuberculosis infection in a patient using  
CC these polypeptides, antibodies or oligonucleotide probes and  
CC primers, for the diagnosis of tuberculosis.  
XX Sequence 100 AA;  
SQ

alignment\_scores:  
Quality: 486.00 Length: 100  
Ratio: 4.860 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 99.000  
alignment\_block:  
US-09-462-480-4 x AAW64339 ..

Align seg 1/1 to: AAW64339 from: 1 to: 100

1 ATGCGAGATGAAGACCGATCGCGCTACCTCGGCGAGGAGCGGTAA 50  
|||||  
1 MetAlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAs 17  
51 TTTTCGAGCGATCTCCGCGACCTGAAACCCAGATCGACCGAGTGGAGT 100  
|||||  
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGlu 34  
101 CGAGCGGAGGTTCTGTCAGGCGCAGTGGCGCGCGCGCGGCGGACGGCC 150  
|||||  
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50  
151 GCCCAGGCGCGGTGGTGGCTTCCAAAGACGACGACCAATAGCAGAAGCA 200  
|||||  
51 AlaGlnAlaAlaValAlaArgPheGlnGlnAlaAlaAsnLysGlnLysG 67  
201 GGAACCTCGACGAGATCTCGACGAATATTCGTACGCGCGCGCTCCAATACT 250  
|||||  
67 nGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyr 84  
251 CGAGGCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 300  
|||||  
84 erArgAlaAspGluGlnGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 100

seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAV39136  
seq\_documentation\_block:  
ID AAY39136 standard; Protein; 100 AA.  
XX  
AC AAY39136;  
XX  
XX  
DT 05-NOV-1999 (first entry)  
XX  
DE M. tuberculosis antigen Tb38-IN amino acid sequence.

XX  
KW Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;  
KW immunotherapy; diagnosis; immunisation; vaccine; infection;  
KW Immune response; skin test.  
XX  
OS Mycobacterium tuberculosis.  
XX  
PN W09942076-A2.  
XX  
XX  
PD 26-AUG-1999.  
XX  
PF 17-FEB-1999; 99NO-US03268.  
XX  
XX 05-MAY-1998; 98US-0072967.

PR 18-FEB-1998; 98US-0025197.  
XX  
XX (CORI-) CORIXA CORP.  
PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;  
PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;  
XX WPI; 1999-527409/44.  
XX  
XX New antigens from Mycobacterium tuberculosis useful in diagnostic  
PT skin tests and protective or therapeutic vaccines or compositions  
XX  
XX Example 3; Page 133-134; 299pp; English.  
XX

CC The present invention describes polypeptides comprising an immunogenic  
CC part of a Mycobacterium tuberculosis antigen (Ag). Also described  
CC are vaccines and fusion protein containing M. tuberculosis Ag's.  
CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and  
CC other polypeptides fragments, can be used in pharmaceutical compositions  
CC or vaccines to generate a protective or therapeutic immune response to  
CC M. tuberculosis and as reagents in skin tests for diagnosis of  
CC tuberculosis. Ag can induce proliferation of, or cytokine secretion  
CC by, T, B or natural killer cells and/or macrophages in  
CC tuberculosis-immune subjects. AAZ19249 to AAZ19460 and AAY39083 to  
CC AAY39225 are used in the exemplification of the present invention.  
XX  
SQ Sequence 100 AA;

alignment\_scores:  
Quality: 486.00 Length: 100  
Ratio: 4.860 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 99.000  
alignment\_block:  
US-09-462-480-4 x AAY39136 ..

Align seg 1/1 to: AAY39136 from: 1 to: 100

1 ATGCGAGATGAAGACCGATCGCGCTACCTCGGCGAGGAGCGGTAA 50  
|||||  
1 MetAlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAs 17  
51 TTTTCGAGCGATCTCCGCGACCTGAAACCCAGATCGACCGAGTGGAGT 100  
|||||  
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGlu 34  
101 CGAGGCGCGAGGTTCTGTCAGGCGCAGTGGCGCGCGCGCGGCGGACGGCC 150  
|||||  
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50  
151 GCCCAGGCGCGGTGGTGGCTTCCAAAGACGACGACCAATAGCAGAAGCA 200  
|||||  
51 AlaGlnAlaAlaValAlaArgPheGlnGlnAlaAlaAsnLysGlnLysG 67  
201 GGAACCTCGACGAGATCTCGACGAATATTCGTACGCGCGCGCTCCAATACT 250  
|||||  
67 nGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyr 84  
251 CGAGGCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 300  
|||||  
84 erArgAlaAspGluGlnGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 100

seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAV38993  
seq\_documentation\_block:  
ID AAY38993 standard; Protein; 100 AA.  
XX  
AC AAY38993;  
XX  
XX  
DT 05-NOV-1999 (first entry)  
XX  
DE M. tuberculosis recombinant antigen protein Tb38-IN.

```
XX Antigen; diagnosis; detection; infection; antibody; immunisation;
KW vaccine; immunity.
XX Mycobacterium tuberculosis.
XX WO9942118-A2.
XX 26-AUG-1999.
XX 17-FEB-1999; 99WO-US03265.
XX 05-MAY-1998; 98US-0072596.
XX 18-FEB-1998; 98US-0024753.
XX (CORI-) CORIXA CORP.
XX Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
PI Lodes MJ, Reed SG, Skeiky YAW, Iwardzik DR, Vedwick TS;
XX WPI; 1999-527416/44.
XX New polypeptide comprising antigenic portions of M. tuberculosis
XX Example 3; Page 179; 323pp; English.
XX This invention describes novel recombinant antigens and their encoding
XX nucleic acids derived from Mycobacterium tuberculosis. The novel
XX polypeptides are useful for detecting M. tuberculosis infection in a
XX biological sample by detecting antibodies which bind with the
XX polypeptides, and are useful as vaccines for immunizing against
XX M. tuberculosis infection. The new detection methods are needed as
XX current vaccination strategies do not provide 100% immunity.
XX Sequence 100 AA;
XX
alignment_scores:
  Quality: 486.00      Length: 100
  Ratio: 4.860        Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 99.000
alignment_block:
US-09-462-480-4 x AAY38993
Align seg 1/1 to: AAY38993 from: 1 to: 100
1 ATGCAGAGATGAAGACCGATGCCGCTACCTCGGCGAGGAGCGGTAA 50
1 MetAlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAs 17
51 TTTCGAGCGGATCTCCGGCGACCTGAAACCCAGATCGACCGGTGGAGT 100
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGlu 34
101 CGACGGCAGGTTCGTTGCGAGCGCGAGTGGCGCGCGCGCGGAGCGGCC 150
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50
151 GCCCAGCGCGGTTGGTGGCTTCCAAAGACGACCCCAATAGCAGCA 200
51 AlaGlnAlaAlaValAlaArgPheGlnGluAlaAlaAsnLysGlnLysG 67
201 GGAACCTCGACGAGATCTCGAGCAATATTCGTACGGCGCGGTCCCAAT 250
67 nGluLeuAspLysIleSerThrAsnIleArgGlnAlaGlyValGlnTrp 84
251 CGAGGGCGGACGAGGACGAGCAGCGCGGTGTCTCGCAAAATGGGCTTC 300
84 erArgAlaAspGluGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 100
seq_name: /SDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.AAB35218
```

```
seq_documentation_block:
ID AAB35218 standard; Protein; 100 AA.
XX AAB35218;
XX 24-APR-2001 (first entry)
XX M tuberculosis RV3874 protein.
XX Tuberculosis; TB; vaccine; esat-6 gene family; RV0287; RV1036C;
KW RV1037C; RV2346C; RV2348C; RV2653C; RV3020C; RV3444C;
KW RV3445C; RV3890C; RV3891C; RV3904C; RV3905C.
XX Mycobacterium tuberculosis.
XX WO200104151-A2.
XX 18-JAN-2001.
XX 13-JUL-2000; 2000WO-DK00398.
XX 13-JUL-1999; 99DK-0001020.
XX 15-JUL-1999; 99US-0144011.
XX (STAT-) STATENS SERUM INST.
XX Andersen P, Skjot R;
PI WPI; 2001-091923/10.
XX New polypeptide encoded by a member of the esat-6-gene family for
XX immunizing against and diagnosis of tuberculosis -
XX Example 2; Page 65; 80pp; English.
XX The present invention provides the protein and coding sequences for
XX members of the esat-6 gene family from Mycobacterium tuberculosis. These
XX proteins include RV0287, RV1036C, RV1037C, RV2346C, RV2348C, RV2653C,
XX RV2654C, RV3020C, RV3444C, RV3445C, RV3890C, RV3891C, RV3904C and
XX RV3905C. These can be used to produce vaccines against, and in the
XX diagnosis of, tuberculosis (TB) infection. The present sequence is one of
XX the proteins of the invention.
XX Sequence 100 AA;
XX
alignment_scores:
  Quality: 486.00      Length: 100
  Ratio: 4.860        Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 99.000
alignment_block:
US-09-462-480-4 x AAB35218
Align seg 1/1 to: AAB35218 from: 1 to: 100
1 ATGCAGAGATGAAGACCGATGCCGCTACCTCGGCGAGGAGCGGTAA 50
1 MetAlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAs 17
51 TTTCGAGCGGATCTCCGGCGACCTGAAACCCAGATCGACCGGTGGAGT 100
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGlu 34
101 CGACGGCAGGTTCGTTGCGAGCGCGAGTGGCGCGCGCGGAGCGGCC 150
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50
151 GCCCAGCGCGGTTGGTGGCTTCCAAAGACGACCCCAATAGCAGCA 200
51 AlaGlnAlaAlaValAlaArgPheGlnGluAlaAlaAsnLysGlnLysG 67
201 GGAACCTCGACGAGATCTCGAGCAATATTCGTACGGCGCGGTCCCAAT 250
```



|||||  
67 ngluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyrS 84  
251 CGAGGGCGGACGAGCAGCAGCGCGTCTCTCGCAATGGCTTC 300  
84 erArgAlaAspGluGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 100

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:AAB19845

seq\_documentation\_block:  
ID\_ AAB19845 standard; Protein; 100 AA.

AC AAB19845;

DT 05-MAR-2001 (first entry)

DE Mycobacterium tuberculosis protein MTBN4.

XX MTBN4; tuberculosis; BCG; vaccine; infection; diagnosis.

XX Mycobacterium tuberculosis.

XX WO2000066157-A1.

XX 09-NOV-2000.

XX 04-MAY-2000; 2000WO-US12257.

XX 04-MAY-1999; 99US-0132505.

XX (PUBL-) PUBLIC HEALTH RES INST NEW YORK.

XX Gennaro ML;

XX WPI; 2001-007153/01.

XX N-PSDB; AAA89038.

XX Novel polypeptide encoded by open reading frames present in  
PT Mycobacterium tuberculosis genome and not by the BCG strain of M.  
PT bovis, useful as vaccine and for diagnosing tuberculosis infection  
XX  
PS Claim 11; Fig 1; 35pp; English.

XX The present sequence is that of the Mycobacterium tuberculosis  
CC MTBN4 protein. This is 1 of 8 proteins, i.e. MTBN1-8 (see  
CC AAB19842-49), encoded by 8 open reading frames (see AAA89035-42)  
CC identified as being present in the genome of M. tuberculosis but  
CC absent from the genome of the BCG strain of Mycobacterium bovis.  
CC MTBN1-8 represent reagents that are useful in discriminating between  
CC M. tuberculosis and BCG and, in particular, for diagnostic methods  
CC which discriminate between exposure of a subject to M. tuberculosis  
CC and vaccination with BCG. The invention features these MTBN  
CC polypeptides, functional fragments of them, DNA encoding them,  
CC vectors, transformed cells, and diagnostic, therapeutic, and  
CC prophylactic (vaccine) methods, including genetic vaccination  
XX methods.

XX Sequence 100 AA;

alignment\_scores:  
Quality: 486.00 Length: 100  
Ratio: 4.860 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 99.000

alignment\_block:  
US-09-462-480-4 x AAB19845 ..

Align seg 1/1 to: AAB19845 from: 1 to: 100

1 ATGGCAGATGACACGATCGCGTACCTCGGCGAGGAGGAGGTAA 50  
|||||  
1 MetAlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAs 17

51 TTTCGAGCGGATCTCCGCGACCTGAAAAACCCAGATCGACCGGTGGAGT 100  
|||||  
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGluS 34  
101 CGAGCGCAGGTTCTTCGAGGGCCAGTGGCGCGCGCGGGGACGGCC 150  
|||||  
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50  
151 GCCCAGCGCGGTGGTGGCTTCCAAAGACGACGCAATAGCAGAAGCA 200  
|||||  
51 AlaGlnAlaAlaValAlaArgPheGlnGluAlaAlaAsnLysGlnLysG 67  
201 GGAATCTGACGAGATCTCGACGAATATTCTCAGCGCGCGCTCCAATAC 250  
|||||  
67 nGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyrS 84  
251 CGAGGGCGGACGAGCAGCAGCGCGTCTCTCGCAATGGCTTC 300  
|||||  
84 erArgAlaAspGluGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 100

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW81746

seq\_documentation\_block:

ID\_ AAW81746 standard; Protein; 802 AA.

XX AAW81746;

XX 27-JAN-1999 (first entry)

XX M. tuberculosis fusion protein Tbf-2.

XX Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;  
XX vaccine; pharmaceutical; infection; diagnosis.

XX Synthetic.

XX Mycobacterium tuberculosis.

XX WO9816646-A2.

XX 23-APR-1998.

XX 07-OCT-1997; 97WO-US18293.

XX 13-MAR-1997; 97US-0818112.

XX 11-OCT-1996; 96US-0730510.

XX (CORI-) CORIXA CORP.

XX Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;  
PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;

XX WPI; 1998-261042/23.

XX N-PSDB; AAV64567.

XX Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used  
PT to develop products for the detection of M. tuberculosis infection  
PT and for diagnosis, treatment and prevention of tuberculosis

XX Disclosure; Page 208-211; 230pp; English.

XX This sequence represents the fusion protein Tbf-2 which is composed of  
CC immunogenic polypeptides from Mycobacterium tuberculosis (MT). This  
CC protein is used in a method for inducing protective immunity against  
CC tuberculosis (TB). This sequence can be formulated into vaccines  
CC and/or pharmaceutical compositions for immunising against  
XX M. tuberculosis infection or may be used for the diagnosis of TB.

XX Sequence 802 AA;

alignment\_scores:  
Quality: 481.00 Length: 99

```
Ratio: 4.859          Gaps: 0
Percent Similarity: 100.000    Percent identity: 98.990
alignment_block:
US-09-462-480-4 x AAW81746  ..
Align seg 1/1 to: AAW81746 from: 1 to: 802
4 GCAGAGATGAAGACCGATGCCGCTACCTCGGGCAGGAGGAGGTAATTT 53
425 AlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAsnPh 441
54 CGAGCGGATCTCCGGCGACCTGAAACCCAGATCGACCGGTGGAGTCGA 103
441 eGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGluSerT 458
104 CGGCGAGTTCTGTCAGGCGCAGTGGCGCGCGCGGCGGAGCGCGCC 153
458 hrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAlaAla 474
154 CAGCGCGGTGGTGGCTTCCAGAACGACGCAATTAAGCAGACGAGGA 203
475 GlnAlaAlaValValArgPheGlnGluAlaAlaAsnLysGlnLysGlnG 491
204 ACTCGACGAGATCTCGACGAATATTCGTACGCGCGCGCTCCAATCT 253
491 uLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyrSe 508
254 GGGCGGACGAGCAGCAGCAGCGCGCTGCTCGCAATGGGCTTC 300
508 rgAlaAspGluGlnGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 523
seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW64379
seq_documentation_block:
ID AAW64379 standard; Protein; 802 AA.
XX AAW64379;
XX
XX 09-NOV-1998 (first entry)
XX
XX Mycobacterium antigen Tbf2 protein fusion.
XX
XX Tuberculosis; infection; diagnosis; 38 kDa antigen; Tbra3; DPEP;
XX TB38-1; Tbf-2.
XX
XX Mycobacterium tuberculosis.
XX Synthetic.
XX
XX WO9816645-A2.
XX
XX 23-APR-1998.
XX
XX 07-OCT-1997; 97WO-US18214.
XX
XX 13-MAR-1997; 97US-0818111.
XX 11-OCT-1996; 96US-0729622.
XX
XX (CORI-) CORIXA CORP.
XX
XX Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
XX Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX
XX WPI; 1998-251292/22.
XX N-PSDB; AAV55801.
XX
XX New isolated Mycobacterium tuberculosis polypeptides and DNA - used
XX PT to develop products for the detection of M. tuberculosis infection
XX and diagnosis of tuberculosis
XX
XX Example 7; Page 223-226; 250pp; English.
XX
XX This polypeptide comprises a fusion protein, designated Tbf-2,
```

```
CC composed of Mycobacterium tuberculosis antigens Tbra3 (see AAW64295),
CC 38 kDa antigen (see AAW64364), Tbf38-1 (see AAW64321) and DPEP (see
CC AAW64322). It was produced by PCR amplification (see AAW4450-57) of
CC the appropriate antigen DNA sequences, cloning into an expression
CC vector, and expression in E. coli. Tbf-2 can be used for
CC serodiagnosis of tuberculosis. The invention relates to
CC compositions and methods for diagnosing tuberculosis. It provides
CC polypeptides (see AAW64291-W64379) comprising antigenic or
CC immunogenic portions of M. tuberculosis antigens, or fusion proteins,
CC DNA sequences encoding such polypeptides, recombinant expression
CC vectors and host cells. Also claimed are methods and diagnostic
CC kits for detecting M. tuberculosis infection in a patient.
XX
SQ Sequence 802 AA;
alignment_scores:
Quality: 481.00 Length: 99
Ratio: 4.859 Gaps: 0
Percent Similarity: 100.000 Percent identity: 98.990
alignment_block:
US-09-462-480-4 x AAW64379  ..
Align seg 1/1 to: AAW64379 from: 1 to: 802
4 GCAGAGATGAAGACCGATGCCGCTACCTCGGGCAGGAGGAGGTAATTT 53
425 AlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAsnPh 441
54 CGAGCGGATCTCCGGCGACCTGAAACCCAGATCGACCGGTGGAGTCGA 103
441 eGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGluSerT 458
104 CGGCGAGTTCTGTCAGGCGCAGTGGCGCGCGCGGCGGAGCGCGCC 153
458 hrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAlaAla 474
154 CAGCGCGGTGGTGGCTTCCAGAACGACGCAATTAAGCAGACGAGGA 203
475 GlnAlaAlaValValArgPheGlnGluAlaAlaAsnLysGlnLysGlnG 491
204 ACTCGACGAGATCTCGACGAATATTCGTACGCGCGCGCTCCAATCT 253
491 uLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyrSe 508
254 GGGCGGACGAGCAGCAGCAGCGCGCTGCTCGCAATGGGCTTC 300
508 rgAlaAspGluGlnGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 523
seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AA132063
seq_documentation_block:
ID AAY32063 standard; Protein; 802 AA.
XX AAY32063;
XX
XX 17-JAN-2000 (first entry)
XX
XX Mycobacterium tuberculosis antigen fusion protein Tbf-2.
XX
XX Tuberculosis; antigen; fusion protein; Tbf-2; Tbra3; 38kD; Tb38-1;
XX DPEP; diagnosis; therapy; vaccine; immunogen.
XX
XX Mycobacterium tuberculosis.
XX
XX WO9951748-A2.
XX
XX 14-OCT-1999.
XX
XX 07-APR-1999; 99WO-US07717.
XX
XX 07-APR-1998; 98US-0056556.
XX
```

PR 30-DEC-1998; 98US-0223040.  
XX (CORI-) CORIXA CORP.  
XX Skeiky YAW, Alderson M, Campos-Neto A;  
PI WPI; 1999-601610/51.  
XX DR N-PSDB; AA220198.  
XX New fusion proteins useful for diagnosis, prevention and treatment of  
PT tuberculosis -  
PT Claim 1; Fig 5G-J; 83pp; English.  
XX  
XX This sequence represents a recombinant Mycobacterium tuberculosis  
CC tetra-antigen fusion protein, termed Tbf-2, composed of the antigens  
CC Tba33, 39kb, Tb38-1 and DPEP. The fusion protein is expressed in  
CC host cells using a vector carrying a polynucleotide (see AA220198)  
CC comprising the 4 coding sequences. The invention provides fusion  
CC proteins (see AA22059-71) containing at least 2 M. tuberculosis  
CC antigens. The new fusion proteins and polynucleotides encoding  
CC them are useful as vaccines for preventing tuberculosis (claimed),  
CC for diagnosis (via in vitro assays or intradermal skin tests for  
CC detection of anti-M. tuberculosis antibodies), monitoring of  
CC disease progression, and treatment of tuberculosis. They are more  
CC effective immunogens than mixtures of the individual protein  
CC components.  
XX  
SQ Sequence 802 AA;

alignment\_scores:  
Quality: 481.00 Length: 99  
Ratio: 4.859 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 98.990  
alignment\_block:  
US-09-462-480-4 x AAY32063 ..  
Align seg 1/1 to: AAY32063 from: 1 to: 802

4 GCAGAGATGAGACCGATGCGCTACCTCGGCGAGGAGCGAGTAATTT 53  
|||||  
425 AlagluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAsnPh 441  
54 CGAGCGGATCTCGGCGGACCTGAAACCCAGATCGACGAGTGAGTCGA 103  
|||||  
441 eGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGluSert 458  
104 CGGAGGTTCTGTTGACGGGCGAGTGGCGCGCGGGGACGCGCGCC 153  
|||||  
458 hrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAlaAla 474  
154 CAGGCCCGGTGGTGGCGCTTCCAAAGAGACGCCAATAAGCAGAGCAGGA 203  
|||||  
475 GlnAlaAlaValValArgPheGlnGlnAlaAlaAsnLysGlnLysGlnGl 491  
204 ACTCGAGAGATCTCGACGAATATTCGTACGGCGGCGGTCCCAATACTCGA 253  
|||||  
491 uLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTrpSerA 508  
254 GGGCCGACGAGGAGCAGCAGCAGCGCTGCTCGCAATGGGCTTC 300  
|||||  
508 rgAlaAspGluGluGlnGlnAlaLeuSerSerGlnMetGlyPhe 523

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT: AAY39224

seq\_documentation\_block:

ID AAY39224 standard; Protein; 802 AA.

XX  
AC AAY39224;

XX 05-NOV-1999 (first entry)

XX M. tuberculosis fusion protein Tbf-6 amino acid sequence.  
DE  
XX Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;  
KW immunotherapy; diagnosis; immunisation; vaccine; infection;  
KW immune response; skin test.  
XX  
OS Synthetic.  
OS Mycobacterium tuberculosis.  
XX  
PN WO9942076-A2.  
XX  
PD 26-AUG-1999.  
XX  
PF 17-FEB-1999; 99WO-US03268.  
XX  
PR 05-MAY-1998; 98US-0072967.  
PR 18-FEB-1998; 98US-0025197.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;  
PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;  
DR WPI; 1999-527409/44.  
DR N-PSDB; AA219457.  
XX  
PT New antigens from Mycobacterium tuberculosis useful in diagnostic  
PT skin tests and protective or therapeutic vaccines or compositions  
XX  
PS Claim 37; Page 271-273; 299pp; English.  
XX  
CC The present invention describes polypeptides comprising an immunogenic  
CC part of a Mycobacterium tuberculosis antigen (Ag). Also described  
CC are vaccines and fusion protein containing M. tuberculosis Ag's.  
CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and  
CC other polypeptides fragments, can be used in pharmaceutical compositions  
CC or vaccines to generate a protective or therapeutic immune response to  
CC M. tuberculosis and as reagents in skin tests for diagnosis of  
CC tuberculosis. Ag can induce proliferation of, or cytokine secretion  
CC by, T, B or natural killer cells and/or macrophages in  
CC tuberculosis-immune subjects. AA219457 and AAY39083 to  
CC AAY39225 are used in the exemplification of the present invention.  
XX  
SQ Sequence 802 AA;

alignment\_scores:  
Quality: 481.00 Length: 99  
Ratio: 4.859 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 98.990  
alignment\_block:  
US-09-462-480-4 x AAY39224 ..  
Align seg 1/1 to: AAY39224 from: 1 to: 802

4 GCAGAGATGAGACCGATGCGCTACCTCGGCGAGGAGCGAGTAATTT 53  
|||||  
425 AlagluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAsnPh 441  
54 CGAGCGGATCTCGGCGGACCTGAAACCCAGATCGACGAGTGAGTCGA 103  
|||||  
441 eGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGluSert 458  
104 CGGAGGTTCTGTTGACGGGCGAGTGGCGCGCGGGGACGCGCGCC 153  
|||||  
458 hrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAlaAla 474  
154 CAGGCCCGGTGGTGGCGCTTCCAAAGAGACGCCAATAAGCAGAGCAGGA 203  
|||||  
475 GlnAlaAlaValValArgPheGlnGlnAlaAlaAsnLysGlnLysGlnGl 491

Align seg 1/1 to: AAY39176 from: 1 to: 802

alignment\_scores:

Quality: 481.00 Length: 99  
Ratio: 4.859 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 98.990  
alignment\_block:  
US-09-462-480-4 x AAY39081 ..

Align seg 1/1 to: AAY39081 from: 1 to: 802

4 GCAGAGATGAAGACCGCGCTACCGCTCGGCGAGGAGCGAGTAATTT 53  
|||||  
425 AlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAsnPh 441  
54 CGACCGGATCTCCGCGCACCTGAAACCCAGATCGACCGAGTGAGTCGA 103  
|||||  
441 eGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGluSert 458  
104 CGGCGAGTTCGTCAGCGCCAGTGGCGCGCGCGCGCGCGCGCGCGCC 153  
|||||  
458 hrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAlaAla 474  
154 CAGCGCGGTGGTGGCTTCCAAAGACGACCAATAAGCAGAAGCAGGA 203  
|||||  
475 GlnAlaAlaValAlaArgPheGlnGluAlaAlaAsnLysGlnLysGlnG 491  
204 ACTCGACGAGATCTCGACGATATTCGTCAGCGCGCGCTCCCAATPACTCGA 253  
|||||  
491 uLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTrpSerA 508  
254 GGGCGCGAGGAGCAGCAGCGCGCTGCTCGCAAAATGGGCTTC 300  
|||||  
508 rgAlaAspGluGlnGlnGlnAlaLeuSerGlnMetGlyPhe 523

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT: AAY39033

seq\_documentation\_block:

ID AAY39033 standard; Protein; 802 AA.  
XX  
AC AAY39033;  
XX  
XX  
DT 05-NOV-1999 (first entry)  
XX  
DE M. tuberculosis fusion protein Tbf-2.  
XX  
KW Antigen; diagnosis; detection; infection; antibody; immunisation;  
KW vaccine; immunity.  
XX  
OS Mycobacterium tuberculosis.  
XX  
PN W09942118-A2.  
XX  
PD 26-AUG-1999.  
XX  
PF 17-FEB-1999; 99NO-US03265.  
XX  
PR 05-MAY-1998; 98US-0072596.  
PR 18-FEB-1998; 98US-0024753.  
XX  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;  
PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedwick TS;  
DR WPI: 1999-527416/44.  
DR N-PSDB; AA219156.  
XX  
XX New polypeptide comprising antigenic portions of M. tuberculosis  
XX  
XX Example 10; Page 251-253; 323pp; English.  
XX  
XX This invention describes novel recombinant antigens and their encoding  
CC nucleic acids derived from Mycobacterium tuberculosis. The novel  
CC polypeptides are useful for detecting M. tuberculosis infection in a

CC biological sample by detecting antibodies which bind with the  
CC polypeptides, and are useful as vaccines for immunizing against  
CC M. tuberculosis infection. The new detection methods are needed as  
CC current vaccination strategies do not provide 100% immunity.  
XX  
SQ Sequence 802 AA;

alignment\_scores:

Quality: 481.00 Length: 99  
Ratio: 4.859 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 98.990

alignment\_block:

US-09-462-480-4 x AAY39033 ..

Align seg 1/1 to: AAY39033 from: 1 to: 802

4 GCAGAGATGAAGACCGATGCGCTACCGCTCGGCGAGGAGCGAGTAATTT 53  
|||||  
425 AlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAsnPh 441  
54 CGACCGGATCTCCGCGCACCTGAAACCCAGATCGACCGAGTGAGTCGA 103  
|||||  
441 eGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGluSert 458  
104 CGGCGAGTTCGTCAGCGCCAGTGGCGCGCGCGCGCGCGCGCGCC 153  
|||||  
458 hrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAlaAla 474  
154 CAGCGCGGTGGTGGCTTCCAAAGACGACCAATAAGCAGAAGCAGGA 203  
|||||  
475 GlnAlaAlaValAlaArgPheGlnGluAlaAlaAsnLysGlnLysGlnG 491  
204 ACTCGACGAGATCTCGACGATATTCGTCAGCGCGCGCTCCCAATPACTCGA 253  
|||||  
491 uLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTrpSerA 508  
254 GGGCGCGAGGAGCAGCAGCGCGCTGCTCGCAAAATGGGCTTC 300  
|||||  
508 rgAlaAspGluGlnGlnGlnAlaLeuSerGlnMetGlyPhe 523

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT: AAU01901

seq\_documentation\_block:

ID AAU01901 standard; Protein; 983 AA.  
XX  
XX AAU01901;  
XX  
DT 29-AUG-2001 (first entry)  
XX  
DE M. tuberculosis Tbf15 fusion protein.  
XX  
KW Tbf15; antigen; vaccine; tuberculosis; AIDS; His tag; Ra3; 38kD;  
KW 38-1; FL TbfH4; acquired immunodeficiency disease.  
XX  
XX  
OS Synthetic.  
OS Mycobacterium tuberculosis.  
XX  
XX  
FH Key Location/Qualifiers  
FH Binding-site 3..8  
FT /label= Histidine tag  
FT /note= "Nickel chelating region used for purifying  
FT the fusion protein"  
FT 9..74  
FT Region /label= Ra3\_region  
FT Region /label= 38kD\_region  
FT Region /label= 38-1\_region  
FT Region /label= 525..983  
FT Region /label= FL\_TbfH4\_region  
XX

PN WO200124820-A1.  
 XX 12-APR-2001.  
 XX 10-OCT-2000; 2000WO-US28095.  
 XX 07-OCT-1999; 99US-0158338.  
 PR 07-OCT-1999; 99US-0158425.  
 XX (CORI-) CORIXA CORP.  
 XX  
 XX Skeiky Y, Reed S, Houghton RL, McNeill PD, Dillon DC, Lodes ML;  
 PI WPI; 2001-290576/30.  
 XX N-PSDB; AAS03795.  
 DR  
 XX  
 PT Vaccinating against Mycobacteria infections in mammals using fusion  
 PT proteins comprising combinations of heterologous antigens -  
 XX  
 PS Claim 6; Fig 4; 168pp; English.  
 XX  
 XX The sequence represents Mycobacterium fusion protein antigen Tbf15  
 CC consisting of a His tag for purification, antigens Ra3, 38KD, 38-1  
 CC and Fl-TbH4 (full-length Tbf4). Compositions comprising at least 2  
 CC heterologous antigens, as a fusion protein, and vectors expressing the  
 CC fusion proteins are used as vaccines to prophylactically immunise mammals  
 CC (especially humans) against infection by Mycobacteria. The compositions  
 CC contain at least 2 heterologous antigens that increase the serological  
 CC sensitivity of individuals infected with tuberculosis, a disease  
 CC frequently affecting patients with acquired immunodeficiency disease,  
 CC AIDS.  
 XX  
 XX Sequence 983 AA;  
 SQ  
 alignment\_scores:  
 Quality: 481.00 Length: 99  
 Ratio: 4.859 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 98.990  
 alignment\_block:  
 US-09-462-480-4 x AAU01901 ..  
 Align seg 1/1 to: AAU01901 from: 1 to: 983  
 4 GCAGAGATGAAGACCGATGCGCTACCTCGGCGAGGAGGAGGAGTAATTT 53  
 426 AlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAsnPh 442  
 54 CGAGCGGATCTCCGGCGACCTGAACCCAGATCGACCGAGTGGAGTCGA 103  
 442 eGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGluSert 459  
 104 CGGCGAGTTCGTCAGGCGCAGTGGCGCGCGCGCGCGCGCGCGCGCC 153  
 459 hrAlaGlySerLeuGlnGlyGlnIleArgGlyAlaAlaGlyThrAlaAla 475  
 154 CAGCGCGCGGTGTGCGCTTCCAAAGACGCCAATAAGCAGAACGACGAGA 203  
 476 GlnAlaAlaValValArgPheGlnGluAlaAlaAsnLysGlnLysGlnG 492  
 204 ACTCGACGAGATCTCGACGAATATTCGTACGCGCGCGGTCCAAATCTCGA 253  
 492 uLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyrSerA 509  
 254 GGGCGGACGAGGACGAGCGCGCTGTCTCGCAAAATGGGCTTC 300  
 509 rgAlaAspGluGluGlnGlnGlnAlaLeuSerSerGlnMetGlyphe 524

OM of: US-09-462-480-4 to: Issued\_Patents\_AA:\* out\_format : pfs  
Date: Jul 22, 2002 1:24 AM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US09462480/runat\_18072002\_164418\_19425/app\_query.fasta\_1.2850  
-DB=Issued\_Patents\_AA -QEMT=fastan -SUFFIX=ra1 -GAPOP=12 000  
-CAPEXT=4 000 -MINWATCH=0 100 -LOOCL=0 000 -LOOPEXT=0 000  
-GAPOP=4 500 -GAPEXT=0 050 -XGAPOP=10 000 -XGAPEXT=0 500  
-FGAPOP=6 000 -FGAPEXT=7 000 -YGAPOP=10 000 -YGAPEXT=0 500  
-DELOP=6 000 -DELEXT=7 000 -START=1 -MATRIX=blomsum62  
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFWT=pfs  
-NORM=ext -HEARSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09462480 -CGN1\_1\_18 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-462-480-4  
Query length: 302  
Database: Issued\_Patents\_AA\*  
Database sequences: 231628  
Database length: 24425594  
Search time (sec): 53.850000

score\_list:

Sequence	Strd Orig	ZScore	Escore	Len	Documentation
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-08-818-112-115		486.00	929.11	2.9e-44	100
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-08-818-111-110		486.00	929.11	2.9e-44	100
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-09-056-556-115		486.00	929.11	2.9e-44	100
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-09-056-556-214		481.00	898.62	1.8e-43	802
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-08-818-112-88		462.00	883.56	1.1e-41	95
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-08-818-111-89		462.00	883.56	1.1e-41	95
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-09-056-556-88		462.00	883.56	1.1e-41	95
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-08-818-112-117		392.00	750.93	3.0e-34	80
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-08-818-111-112		392.00	750.93	3.0e-34	80
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-09-056-556-117		392.00	750.93	3.0e-34	80
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-08-818-112-98		122.00	243.26	1.7e-05	28
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-09-056-556-98		122.00	243.26	1.7e-05	28
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-08-818-112-96		121.00	241.70	2.1e-05	27
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-09-056-556-96		121.00	241.70	2.1e-05	27
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-08-818-112-95		118.00	235.94	4.4e-05	27
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-09-056-556-95		118.00	235.94	4.4e-05	27
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-08-818-112-97		118.00	235.94	4.4e-05	27
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-09-056-556-97		118.00	235.94	4.4e-05	27
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-08-818-112-93		114.00	227.90	0.0001	28
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-09-056-556-93		114.00	227.90	0.0001	28
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-09-220-528-116		98.50	178.37	0.0095	201
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-09-220-528-26		92.50	165.95	0.0425	220
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-09-220-528-32		92.50	165.20	0.0434	237
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-09-220-528-40		91.00	165.03	0.0582	181
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-09-199-637A-369		90.50	160.24	0.0734	265
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-09-413-814-78		88.50	144.33	0.1696	882
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-08-628-829-4		88.50	138.40	0.2010	1593
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-09-199-637A-319		88.00	165.22	0.1026	100
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-07-626-618A-17		86.00	147.80	0.2477	387
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-08-086-439C-3		86.00	147.80	0.2477	387
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-08-056-051-2		86.00	147.80	0.2477	387
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-07-928-611-18		86.00	147.80	0.2477	387
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-08-333-977-17		86.00	147.80	0.2477	387
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-08-434-153-1		86.00	147.80	0.2477	387
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-08-434-877-3		86.00	147.80	0.2477	387
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-08-449-651-1		86.00	147.80	0.2477	387
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-08-487-811A-18		86.00	147.80	0.2477	387
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-08-833-703-1		86.00	147.80	0.2477	387
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-08-475-742-4		86.00	147.80	0.2477	387
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-09-060-694-18		86.00	147.80	0.2477	387

seq\_name: /cgn2\_6/ptodata/2/iaa/6B\_COMB.ppt:US-08-818-112-115

seq documentation\_block:

Sequence 115, Application US/08818112  
Patent No. 6290969  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Skeiky, Yasir A.W.  
APPLICANT: Dillon, Davin C.  
APPLICANT: Campos-Neto, Antonio  
APPLICANT: Houghton, Raymond  
APPLICANT: Vedvick, Thomas S.  
APPLICANT: Twardzik, Daniel R.  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY  
TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS  
NUMBER OF SEQUENCES: 153  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/818,112  
FILING DATE: 13-MAR-1997  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.411c6  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 115:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 100 amino acids  
TYPE: amino acid  
STRANDEDNESS: linear  
TOPOLOGY: linear  
US-08-818-112-115

alignment\_scores:

Quality: 486.00 Length: 100  
Ratio: 4.860 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 99.000

alignment\_block:

US-09-462-480-4 x US-08-818-112-115 ..

Align seg 1/1 to: US-08-818-112-115 from: 1 to: 100

1 ATGCAGAGATGACACCGATCGCGCTACCTCGGGCAGGAGCGGTAA 50

|||||

1 MetAlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAs 17

|||||

51 TTTCAGCGGATCTCCGCGACCTCAAAACCCAGATCCAGCGGTGGAGT 100

|||||

17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGlu 34

|||||

101 CGACGGCAGGTTCTGTTGAGGCCAGTGGCGCGCGCGGCGGACGCC 150

|||||  
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50  
151 GCCAGGCGCGGTGCTCCAGCAAGCAGCAATAAGCAGAACA 200  
|||||  
51 AlaGlnAlaAlaValValArgPheGlnGlnAlaAlaAsnLysGlnLysG1 67  
201 GGAAGTCCAGCAGATCTCGAGCAATATTCTGTCAGGCGCGTCCCAATACT 250  
|||||  
67 nGluLeuAspGluLeuSerThrAsnIleArgGlnAlaGlyValGlnTyrS 84  
251 CGAGGCGCGCAGCAGCAGCAGCGGCTGCTCCCAATGGCTTC 300  
|||||  
84 erArgAlaAspGluGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 100  
seq\_name: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:US-08-818-111-110

seq\_documentation\_block:  
; Sequence 110, Application US/08818111  
; Patent No. 6338852  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Campos-Neto, Antonia  
; APPLICANT: Houghton, Raymond  
; APPLICANT: Vedwick, Thomas S.  
; APPLICANT: Twardzik, Daniel R.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF  
; NUMBER OF SEQUENCES: 148  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/818,111  
; FILING DATE: 13-MAR-1997  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.417C6  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 110:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 100 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; US-08-818-111-110

alignment\_scores:  
Quality: 486.00 Length: 100  
Ratio: 4.860 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 99.000

alignment\_block:  
US-09-462-480-4 x US-08-818-111-110 ..  
Align seg 1/1 to: US-08-818-111-110 from: 1 to: 100

1 ATGCGAGATGAAGACCGATGCCGCTACCCCTCGGCGAGGAGGAGTAA 50

|||||  
1 MetAlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAs 17  
51 TTTCAGCGGATCTCCGGCGCCTCAAAACCCAGATCGACCAAGTGGAGT 100  
|||||  
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleaspGlnValGluS 34  
101 CGACGCGCAGGTTCGTTGACAGGCCAGTGGCGCGCGCGGGGACGGCC 150  
|||||  
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50  
151 GCCAGGCGCGGTGCTCCAGCAAGCAGCAATAAGCAGAACA 200  
|||||  
51 AlaGlnAlaAlaValValArgPheGlnGlnAlaAlaAsnLysGlnLysG1 67  
201 GGAAGTCCAGCAGATCTCGAGCAATATTCTGTCAGGCGCGTCCCAATACT 250  
|||||  
67 nGluLeuAspGluLeuSerThrAsnIleArgGlnAlaGlyValGlnTyrS 84  
251 CGAGGCGCGCAGCAGCAGCAGCGGCTGCTCCCAATGGCTTC 300  
|||||  
84 erArgAlaAspGluGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 100  
seq\_name: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:US-09-056-556-115

seq\_documentation\_block:  
; Sequence 115, Application US/09056556  
; Patent No. 6350456  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Dillon, Davin C.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND  
; NUMBER OF SEQUENCES: 241  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/056,556  
; FILING DATE: 07-APR-1998  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.457  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 115:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 100 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; US-09-056-556-115

alignment\_scores:  
Quality: 486.00 Length: 100  
Ratio: 4.860 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 99.000

alignment\_block:  
US-09-462-480-4 x US-09-056-556-115 ..



Align seg 1/1 to: US-09-056-556-115 from: 1 to: 100.

```
1 ATGGCAGAGATGAAGACCGATGCCGTACCTCGGCGCAGGAGCGAGTAA 50
1 MetAlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAs 17
51 TTTTCGAGCGATCTCCGGCGACCTGAAACCCAGATCGACCGAGTGGAGT 100
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGluS 34
101 CGACGCGAGTTCTGTCGAGGCGCATGGCGCGCGCGCGGAGCGGCC 150
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50
151 GCCACGCGCGGTGTCGCTTCCAGAGCAGCCCAATAGCAGAGCA 200
51 AlaGlnAlaAlaValValArgPheGlnGluAlaAlaAsnLysGlnLysG 67
201 GGAACTCGACGAGATCTCGACGAATATTCGTACAGCGCGCGTCCAACT 250
67 nGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyrS 84
251 CGAGGCGCGCAGGAGCAGCAGCAGCGCTGTCTCGCAAAATGGCGCTTC 300
84 erArgAlaAspGluGluGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 100
seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pap:US-09-056-556-214
```

seq\_documentation\_block:  
; Sequence 214, Application US/09056556  
; Patent No. 6350456

GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Dillon, Davin C.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND  
; NUMBER OF SEQUENCES: 241  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED AND BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/056,556  
; FILING DATE: 07-APR-1998  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.457  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 214:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 802 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-09-056-556-214

alignment\_scores:  
Quality: 481.00 Length: 99  
Ratio: 4.859 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 98.990  
alignment\_block:  
US-09-462-480-4 x US-09-056-556-214 ..  
Align seg 1/1 to: US-09-056-556-214 from: 1 to: 802  
4 GCAGAGATGAAGACCGATGCCGTACCTCGGCGCAGGAGCGAGTAAATTT 53  
|||||  
425 AlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAsnPh 441  
|||||  
54 CGAGCGGATCTCCGGCGACCTGAAACCCAGATCGACCGAGTGGAGTCCA 103  
|||||  
441 eGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGluSerT 458  
|||||  
104 CGCAGGTTCTGTCGAGGCGCAGTGGCGCGCGCGCGGAGCGGCC 153  
|||||  
458 hrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAlaAla 474  
|||||  
154 CAGCGCGCGGTGTCGCTTCCAGAGCAGCCCAATAGCAGAGCAGGA 203  
|||||  
475 GlnAlaAlaValValArgPheGlnGluAlaAlaAsnLysGlnLysGlnG 491  
|||||  
204 ACTCGACGAGATCTCGACGAATATTCGTACAGCGCGCGTCCAAATCTCA 253  
|||||  
491 uLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyrSerA 508  
|||||  
254 GCGCGCGCAGGAGCAGCAGCAGCGCTGTCTCGCAAAATGGCGCTTC 300  
|||||  
508 rgAlaAspGluGluGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 523  
seq\_name: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pap:US-08-818-112-88

seq\_documentation\_block:  
; Sequence 88, Application US/08818112  
; Patent No. 6290969  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Campos-Neto, Antonio  
; APPLICANT: Houghton, Raymond  
; APPLICANT: Vedvick, Thomas S.  
; APPLICANT: Twardzik, Daniel R.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY  
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS  
; NUMBER OF SEQUENCES: 153  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED AND BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/818,112  
; FILING DATE: 13-MAR-1997  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.411C6  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 88:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 95 amino acids

TRE

```
; CURRENT APPLICATION DATA:
```

```
;
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 88:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 95 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-056-556-88

alignment_scores:
  Quality: 462.00      Length: 95
  Ratio: 4.863        Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 98.947

alignment_block:
US-09-462-480-4 x US-09-056-556-88 ..
Align seg 1/1 to: US-09-056-556-88 from: 1 to: 95

16 ACCGATCCCTACCTCGGCGAGGAGGTAATTTCAGCGGATCTC 65
|||||
1 ThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAsnPheGluArgIleSe 17

66 CGCGCAGCTGAAACCAGATCGACAGGTGGAGTCGACGCGAGTTTCGT 115
|||||
17 rGlyAspLeuLysThrGlnIleAspGlnValGluSerThrAlaGlySerL 34

116 TGCAGGCGCCAGTGGCGCGCGCGGGGACGGCGCCCGCGCGGTG 165
|||||
34 euGlnGlyGlnTrpArgGlyAlaAlaGlyThrAlaAlaGlnAlaAval 50

166 GTGGCTTCCAGACGAGCAGCCATAAGCAGACGAGGAACTCGACGAGAT 215
|||||
51 ValArgPheGlnGluAlaAlaAsnLysGlnLysGlnGluLeuAspGluIl 67

216 CTCGACGAATATTCGTGAGCGCGCGTCCCAATCTCGCAATGGGCTTC 265
|||||
67 eSerThrAsnIleArgGlnAlaGlyValGlnTyrSerArgAlaAspGlu 84

266 AGCAGCAGCAGCGCTCTCTCGCAATGGGCTTC 300
|||||
84 luGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 95

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-08-818-112-117

seq_documentation_block:
; Sequence 117, Application us/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSES: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
```

```
;
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 117:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 80 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-818-112-117

alignment_scores:
  Quality: 392.00      Length: 80
  Ratio: 4.900        Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-462-480-4 x US-08-818-112-117 ..
Align seg 1/1 to: US-08-818-112-117 from: 1 to: 80

61 ATCTCCGGCGACCTGAAACCCAGATCGACCGAGTGGAGTCGACGCGCAGG 110
|||||
1 IleSerGlyAspLeuLysThrGlnIleAspGlnValGluSerThrAlaGl 17

111 TTCTTCAGGCGCAGTGGCGCGCGCGGGGACGGCGCCCGCCAGGCGCG 160
|||||
17 ySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAlaAlaGlnAlaA 34

161 CGGTGGTGGCTTCCAGAGACGCCCAATAGCAGACGAGGAACTCGAC 210
|||||
34 laValValArgPheGlnGluAlaAlaAsnLysGlnLysGlnGluLeuAsp 50

211 GAGATCTCGACGAATATTCGTGAGCGCGGTCCCAATCTCGAGGCGCGCA 260
|||||
51 GluIleSerThrAsnIleArgGlnAlaGlyValGlnTyrSerArgAlaAs 67

261 CGAGGAGCAGCAGCGCGTCTCTCGCAATGGGCTTC 300
|||||
67 pGluGluGlnGlnAlaLeuSerSerGlnMetGlyPhe 80

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-08-818-111-112

seq_documentation_block:
; Sequence 112, Application us/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; TUBERCULOSIS
```

ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/818,111  
FILING DATE: 13-MAR-1997  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.417C6  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 112:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 80 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-818-111-112

alignment\_scores:  
Quality: 392.00 Length: 80  
Ratio: 4.900 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-462-480-4 x US-08-818-111-112 ..

Align seg 1/1 to: US-08-818-111-112 from: 1 to: 80

61 ATCTCCGGCGACCTGAAACCCAGATCGACGAGTGGAGTCGACGGCAGG 110  
|||||  
1 IleSerGlyAspLeuIysThrGlnIleAspGlnValGluSerThrAlaG1 17  
111 TTCGTTCCAGGCGCAGTGGCGCGCGCGGGGACGCGCGCCAGCCG 160  
|||||  
17 YSerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAlaAlaGlnAla 34  
161 CGGTGGTGCCTTCCAAAGACGACCCCAATAGCAGAGCAGGAACCTGCAC 210  
|||||  
34 laValValArgPheGlnGluAlaAlaAsnLysGlnLysGlnLeuAsp 50  
211 GAGATCTCGACGATATTCGTCAGCGCGCGGTCCCAATCTCGAGGCGCA 260  
|||||  
51 GluIleSerThrAsnIleArgGlnAlaGlyValGlnTyrSerArgAlaAs 67  
261 CGAGGACGACGACGCGCTGTCTCGCAATGGGCTTC 300  
|||||  
67 pGluGluGlnGlnAlaLeuSerSerGlnMetGlyPhe 80

seq\_name: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep:US-09-056-556-117

seq\_documentation\_block:

; Sequence 117, Application US/09056556  
; Patent No. 6350456  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Dillon, Davin C.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND  
; NUMBER OF SEQUENCES: 241  
; CORRESPONDENCE ADDRESS:

TREATM

ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/056,556  
FILING DATE: 07-APR-1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.457  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 117:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 80 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-056-556-117

alignment\_scores:  
Quality: 392.00 Length: 80  
Ratio: 4.900 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-462-480-4 x US-09-056-556-117 ..

Align seg 1/1 to: US-09-056-556-117 from: 1 to: 80

61 ATCTCCGGCGACCTGAAACCCAGATCGACGAGTGGAGTCGACGGCAGG 110  
|||||  
1 IleSerGlyAspLeuIysThrGlnIleAspGlnValGluSerThrAlaG1 17  
111 TTCGTTCCAGGCGCAGTGGCGCGCGGGGACGCGCGCCAGCCG 160  
|||||  
17 YSerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAlaAlaGlnAla 34  
161 CGGTGGTGCCTTCCAAAGACGACCCCAATAGCAGAGCAGGAACCTGCAC 210  
|||||  
34 laValValArgPheGlnGluAlaAlaAsnLysGlnLysGlnLeuAsp 50  
211 GAGATCTCGACGATATTCGTCAGCGCGCGGTCCCAATCTCGAGGCGCA 260  
|||||  
51 GluIleSerThrAsnIleArgGlnAlaGlyValGlnTyrSerArgAlaAs 67  
261 CGAGGACGACGACGCGCTGTCTCGCAATGGGCTTC 300  
|||||  
67 pGluGluGlnGlnAlaLeuSerSerGlnMetGlyPhe 80

seq\_name: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep:US-08-818-112-98

seq\_documentation\_block:

; Sequence 98, Application US/08818112  
; Patent No. 6290969  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Campos-Neto, Antonio  
; APPLICANT: Houghton, Raymond  
; APPLICANT: Vedvick, Thomas S.

; APPLICANT: Twardzik, Daniel R.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY  
; NUMBER OF SEQUENCES: 153  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED AND BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/818,112  
; FILING DATE: 13-MAR-1997  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.411C6  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 98:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 28 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-818-112-98

alignment\_scores:  
Quality: 122.00 Length: 25  
Ratio: 4.880 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-462-480-4 x US-08-818-112-98 ..

Align seg 1/1 to: US-08-818-112-98 from: 1 to: 28

226 ATTCGTCAGCGCGCTCCCAATCTCGAGCGCGAGGAGGAGGAGCA 275  
|||||  
4 IleArgGlnAlaGlyValGlnTyrSerArgAlaAspGluGlnGlnG1 20  
276 GCGCTGCTCTCGCAATGGCTTC 300  
|||||  
20 nAlaLeuSerSerGlnMetGlyPhe 28

seq\_name: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep:US-09-056-556-98

seq\_documentation\_block:  
; Sequence 98, Application US/09056556  
; Patent No. 6350456  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Dillon, Davin C.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND  
; NUMBER OF SEQUENCES: 241  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED AND BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/056,556  
; FILING DATE: 07-APR-1998  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.457  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 98:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 28 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-09-056-556-98

alignment\_scores:  
Quality: 122.00 Length: 25  
Ratio: 4.880 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-462-480-4 x US-09-056-556-98 ..

Align seg 1/1 to: US-09-056-556-98 from: 1 to: 28

226 ATTCGTCAGCGCGCTCCCAATCTCGAGCGCGAGGAGGAGGAGCA 275  
|||||  
4 IleArgGlnAlaGlyValGlnTyrSerArgAlaAspGluGlnGlnG1 20  
276 GCGCTGCTCTCGCAATGGCTTC 300  
|||||  
20 nAlaLeuSerSerGlnMetGlyPhe 28

seq\_name: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep:US-08-818-112-96

seq\_documentation\_block:  
; Sequence 96, Application US/08818112  
; Patent No. 6290969  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Campos-Neto, Antonio  
; APPLICANT: Houghton, Raymond  
; APPLICANT: Twardzik, Daniel R.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY  
; NUMBER OF SEQUENCES: 153  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED AND BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/818,112  
; FILING DATE: 13-MAR-1997

```

; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-056-556-96

alignment_scores:
    Quality: 121.00      Length: 27
    Ratio: 4.481        Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 92.593

alignment_block:
    US-09-462-480-4 x US-09-056-556-96 ..
    Align seg 1/1 to: US-09-056-556-96 from: 1 to: 27

133 GCGCGGGGGGACGGCCGCCAGCGCGGTGGTGCCTTCCAAAGAC 182
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1 GlycylGlyGlyThrAlaAlaGlnAlaValArgPheGlnGluAl 17

183 AGCCAATAAGCAGAGCAGGAGCACTCGACGAG 213
|||||:|||||:|||||:|||||:|||||:|||||:|||||
17 alaAsnLysGlnLysGlnGluLeuAspGlu 27

seg_name: /cgn2_6/ptodata/2/iaa/6B_COMB pep:US-08-818-112-95

seq_documentation_block:
; Sequence 95, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 5300 Columbia Center, 701 Fifth Avenue
; City: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 95:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 27 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-818-112-95

alignment_scores:
    Quality: 118.00      Length: 24

```

```

14
      Ratio: 4.917          Gaps: 0
Percent Similarity: 100.000  Percent Identity: 100.000
alignment_block:
US-09-462-480-4 x US-08-818-112-95  ..
Align seg 1/1 to: US-08-818-112-95 from: 1 to: 27
100 TCGACGGCAGGTTCTTGCAGGCCAGTGGCGCGCGCGCGGGGACGGC 149
|||||
4 SerThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAl 20
150 CGCCCAAGCGCGGTGTGCGC 171
|||||
20 aAlaGlnAlaAlaValValArg 27

```

THIS PAGE BLANK (USPTO)



OM of: US-09-462-480-4 to: Pending\_Patents\_AA\_Main:\* out\_format : pfs  
Date: Jul 22, 2002 1:35 AM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:  
-MODEL=frame+ n2p.model -DEV=xlh  
-O=/cgn2\_1/USPTO.spool/US09462480/runat\_18072002\_164419\_19468/app\_query.fasta\_1.2850  
-DB=Pending\_Patents\_AA\_Main -QFMT=fastan -SUFFIX=rapm  
-CAPOP=12.000 -CAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000  
-LOOPEXT=0.000 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000  
-XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000  
-YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1  
-MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOALIGN=200  
-THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -LOCAL  
-OUTFMT=pfs -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09462480 -CGN1\_1\_271 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-462-480-4  
Query length: 302  
Database: Pending\_Patents\_AA\_Main:\*  
Database sequences: 3502263  
Database length: 351980561  
Search time (sec): 542.380000

score\_list:

Sequence	Strd Orig	zScore	Escore	Len	Documentation
/cgn2_6/ptodata/2/paa/US091_COMB.ppt	US-09-116-492-5	492.00	727.11	7.5e-32	100
/cgn2_6/ptodata/2/paa/US094_COMB.ppt	US-09-462-480-5	492.00	727.11	7.5e-32	100
/cgn2_6/ptodata/2/paa/PCTUS_COMB.ppt	PCT-US99-03265-110	486.00	718.32	2.3e-31	100
/cgn2_6/ptodata/2/paa/PCTUS_COMB.ppt	PCT-US99-03265-115	486.00	718.32	2.3e-31	100
/cgn2_6/ptodata/2/paa/US087_COMB.ppt	US-08-729-622-110	486.00	718.32	2.3e-31	100
/cgn2_6/ptodata/2/paa/US087_COMB.ppt	US-08-730-510-115	486.00	718.32	2.3e-31	100
/cgn2_6/ptodata/2/paa/US089_COMB.ppt	US-08-942-341-110	486.00	718.32	2.3e-31	100
/cgn2_6/ptodata/2/paa/US089_COMB.ppt	US-08-942-578-115	486.00	718.32	2.3e-31	100
/cgn2_6/ptodata/2/paa/US090_COMB.ppt	US-09-024-753-110	486.00	718.32	2.3e-31	100
/cgn2_6/ptodata/2/paa/US090_COMB.ppt	US-09-025-197-115	486.00	718.32	2.3e-31	100
/cgn2_6/ptodata/2/paa/US090_COMB.ppt	US-09-072-596-110	486.00	718.32	2.3e-31	100
/cgn2_6/ptodata/2/paa/US090_COMB.ppt	US-09-072-596-115	486.00	718.32	2.3e-31	100
/cgn2_6/ptodata/2/paa/US096_COMB.ppt	US-08-615-947-1	486.00	718.32	2.3e-31	100
/cgn2_6/ptodata/2/paa/US097_COMB.ppt	US-09-724-685-115	486.00	718.32	2.3e-31	100
/cgn2_6/ptodata/2/paa/US100_COMB.ppt	US-10-084-843-115	486.00	718.32	2.3e-31	100
/cgn2_6/ptodata/2/paa/PCTUS_COMB.ppt	PCT-US99-03265-209	481.00	693.34	7.1e-31	802
/cgn2_6/ptodata/2/paa/PCTUS_COMB.ppt	PCT-US99-03265-346	481.00	693.34	7.1e-31	802
/cgn2_6/ptodata/2/paa/PCTUS_COMB.ppt	PCT-US99-03268-214	481.00	693.34	7.1e-31	802
/cgn2_6/ptodata/2/paa/PCTUS_COMB.ppt	PCT-US99-03268-351	481.00	693.34	7.1e-31	802
/cgn2_6/ptodata/2/paa/US089_COMB.ppt	US-08-942-341-210	481.00	693.34	7.1e-31	802
/cgn2_6/ptodata/2/paa/US089_COMB.ppt	US-08-942-578-214	481.00	693.34	7.1e-31	802
/cgn2_6/ptodata/2/paa/US090_COMB.ppt	US-09-024-753-209	481.00	693.34	7.1e-31	802
/cgn2_6/ptodata/2/paa/US090_COMB.ppt	US-09-025-197-214	481.00	693.34	7.1e-31	802
/cgn2_6/ptodata/2/paa/US090_COMB.ppt	US-09-072-596-209	481.00	693.34	7.1e-31	802
/cgn2_6/ptodata/2/paa/US090_COMB.ppt	US-09-072-596-346	481.00	693.34	7.1e-31	802
/cgn2_6/ptodata/2/paa/US090_COMB.ppt	US-09-072-967-214	481.00	693.34	7.1e-31	802
/cgn2_6/ptodata/2/paa/US092_COMB.ppt	US-09-072-967-351	481.00	693.34	7.1e-31	802
/cgn2_6/ptodata/2/paa/US092_COMB.ppt	US-09-287-849-10	481.00	693.34	7.1e-31	802
/cgn2_6/ptodata/2/paa/US100_COMB.ppt	US-10-084-843-214	481.00	693.34	7.1e-31	802
/cgn2_6/ptodata/2/paa/US100_COMB.ppt	US-10-084-843-351	481.00	693.34	7.1e-31	802
/cgn2_6/ptodata/2/paa/PCTUS_COMB.ppt	PCT-US99-03265-89	462.00	683.63	2.1e-29	95
/cgn2_6/ptodata/2/paa/PCTUS_COMB.ppt	PCT-US99-03268-88	462.00	683.63	2.1e-29	95
/cgn2_6/ptodata/2/paa/US086_COMB.ppt	US-08-658-800-88	462.00	683.63	2.1e-29	95
/cgn2_6/ptodata/2/paa/US086_COMB.ppt	US-08-659-683-88	462.00	683.63	2.1e-29	95
/cgn2_6/ptodata/2/paa/US086_COMB.ppt	US-08-680-573-88	462.00	683.63	2.1e-29	95
/cgn2_6/ptodata/2/paa/US086_COMB.ppt	US-08-680-574-88	462.00	683.63	2.1e-29	95
/cgn2_6/ptodata/2/paa/US087_COMB.ppt	US-08-729-622-88	462.00	683.63	2.1e-29	95
/cgn2_6/ptodata/2/paa/US087_COMB.ppt	US-08-730-510-88	462.00	683.63	2.1e-29	95
/cgn2_6/ptodata/2/paa/US089_COMB.ppt	US-08-942-341-88	462.00	683.63	2.1e-29	95
/cgn2_6/ptodata/2/paa/US089_COMB.ppt	US-08-942-578-88	462.00	683.63	2.1e-29	95

/cgn2\_6/ptodata/2/paa/US090\_COMB.ppt: US-09-024-753-89 + 462.00 683.63 2.1e-29  
/cgn2\_6/ptodata/2/paa/US090\_COMB.ppt: US-09-025-197-88 + 462.00 683.63 2.1e-29  
/cgn2\_6/ptodata/2/paa/US090\_COMB.ppt: US-09-072-596-89 + 462.00 683.63 2.1e-29  
/cgn2\_6/ptodata/2/paa/US090\_COMB.ppt: US-09-072-967-88 + 462.00 683.63 2.1e-29  
/cgn2\_6/ptodata/2/paa/US097\_COMB.ppt: US-09-724-685-88 + 462.00 683.63 2.1e-29  
seq\_name: /cgn2\_6/ptodata/2/paa/US091\_COMB.ppt: US-09-116-492-5

seq\_documentation\_block:

; Sequence 5, Application US/09116492  
; GENERAL INFORMATION:  
; APPLICANT: GICQUEL, BRIGITTE  
; APPLICANT: BERTHET, FRANCOIS-XAVIER  
; APPLICANT: ANDERSEN, PETER  
; APPLICANT: RASMUSSEN, PETER B  
; TITLE OF INVENTION: POLYNUCLEOTIDE FUNCTIONALLY CODING FOR THE LHP PROTEIN  
; TITLE OF INVENTION: FROM MYCOBACTERIUM TUBERCULOSIS, ITS BIOLOGICALLY  
; TITLE OF INVENTION: ACTIVE DERIVATIVE FRAGMENTS, AS WELL AS METHODS USING  
; TITLE OF INVENTION: THE SAME  
; FILE REFERENCE: 0660-0137-27X  
; CURRENT APPLICATION NUMBER: US/09/116,492  
; EARLIER APPLICATION NUMBER: 60/052,631  
; EARLIER FILING DATE: 1998-07-16  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 100  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis  
US-09-116-492-5

alignment\_scores:

Quality: 492.00 Length: 100  
Ratio: 4.920 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-462-480-4 x US-09-116-492-5

Align seg 1/1 to: US-09-116-492-5 from: 1 to: 100

1	ATGCGACGATGAACGACGCGCTACCTCGCGGAGGAGGAGTAA	50
1	MetAlaGluMetLysThrAspAlaAlaThrLeuGlyGlnAlaGlyAs	17
51	TTTCGAGCGGATCTCCGCGCACCTGAAACCCAGATCGACGAGTGAGT	100
17	nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGlu	34
101	CGACGCGAGGTTCTGTCAGCGGCGAGTGGCGGCGGCGGCGGCGGCC	150
34	erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla	50
151	GCCAGCGCGCGGTGGTGGCGCTTCCAAAGACAGCAATTAAGCAGAAGCA	200
51	AlaGlnAlaAlaValArgPheGlnGlnAlaAlaAsnLysGlnLysGln	67
201	GGAACCTCGACGAGATCTCGAGCAATATTCGTACGCGCGGCGTCCCAAT	250
67	nGluLeuAspLutIleSerThrAsnIleArgGlnAlaGlyValGlnTyr	84
251	CGAGGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	300
84	erArgAlaAspGluGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe	100

seq\_name: /cgn2\_6/ptodata/2/paa/US094\_COMB.ppt: US-09-462-480-5

seq\_documentation\_block:

; Sequence 5, Application US/09462480  
; GENERAL INFORMATION:  
; APPLICANT: GICQUEL, BRIGITTE

APPLICANT: BERTHET, FRANCIS-XAVIER  
APPLICANT: ANDERSEN, PETER  
APPLICANT: RASMUSSEN, PETER BIRK  
TITLE OF INVENTION: POLYNUCLEOTIDE FUNCTIONALLY CODING FOR THE LHP PROTEIN FROM MYCOBACTERIUM TUBERCULOSIS, ITS BIOLOGICALLY ACTIVE DERIVATIVE FRAGMENTS, AND METHODS FOR IMMUNOTHERAPY  
TITLE OF INVENTION: USING THE SAME  
FILE REFERENCE: 0660-0165-0XPCT  
CURRENT APPLICATION NUMBER: US/09/462,480  
PRIOR FILING DATE: 2000-03-06  
PRIOR APPLICATION NUMBER: PCT/IB98/01091  
PRIOR FILING DATE: 1998-07-16  
PRIOR APPLICATION NUMBER: 60/052,631  
PRIOR FILING DATE: 1997-07-16  
NUMBER OF SEQ ID NOS: 34  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 5  
LENGTH: 100  
TYPE: PRT  
ORGANISM: Mycobacterium tuberculosis  
US-09-462-480-5

alignment\_scores:  
Quality: 492.00 Length: 100  
Ratio: 4.920 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-462-480-4 x US-09-462-480-5  
Align seg 1/1 to: US-09-462-480-5 from: 1 to: 100  
1 ATGGCAGAGATGAAGACCGGATCGCGCTACCGTCCGGCAGGAGGAGGTAA 50  
1 MetAlaGluMetLysThrAspAlaAlaThrLeuGlnGluAlaGlyAs 17  
51 TTTCGAGCGGATCTCCGGGACCTGAAACCCAGATCGACCGAGTGGAGT 100  
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGlu 34  
101 CGACGGCAGGTTCTTCGAGGCGCAGTGGCGCGCGCGGCGGAGCGGCC 150  
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50  
151 GCCAGCGCGCGGTGGTGGCTTCCAGAACGAGCCCAATAAGCAGAGCA 200  
51 AlaGlnAlaAlaValAlaArgPheGlnGluAlaAlaAsnLysGlnLysG 67  
201 GGAACCTGACGAGATCTCGACGAAATATTCGTCAGCGCGCGGTCCAACT 250  
67 nGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyr 84  
251 CGAGGGCGCGGAGGAGCAGCAGCAGCGCTGTCTCGCAAAATGGGCTTC 300  
84 erArgAlaAspGluGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 100  
seq\_name: /cgn2\_6/ptodata/2/paa/PCTUS\_COMB.pep:PCT-US99-03265-110

seq\_documentation\_block:

Sequence 110, Application PC/TUS9903265  
GENERAL INFORMATION:  
APPLICANT: Corixa Corporation  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS  
NUMBER OF SEQUENCES: 350  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds, LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2811  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US99/03265  
FILING DATE: 17-FEB-1999  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 09/024,753  
FILING DATE: 18-FEB-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Laura A. Coruzzi  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 9532-0023-228  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-493-4935  
TELEFAX: 650-493-5556  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 110:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 100 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
PCT-US99-03265-110

alignment\_scores:  
Quality: 486.00 Length: 100  
Ratio: 4.860 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 99.000

alignment\_block:

US-09-462-480-4 x PCT-US99-03265-110  
Align seg 1/1 to: PCT-US99-03265-110 from: 1 to: 100  
1 ATGGCAGAGATGAAGACCGGATCGCGCTACCGTCCGGCAGGAGGAGGTAA 50  
1 MetAlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAs 17  
51 TTTCGAGCGGATCTCCGGGACCTGAAACCCAGATCGACCGAGTGGAGT 100  
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGlu 34  
101 CGACGGCAGGTTCTTCGAGGCGCAGTGGCGCGCGCGGCGGAGCGGCC 150  
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50  
151 GCCAGCGCGCGGTGGTGGCTTCCAGAACGAGCCCAATAAGCAGAGCA 200  
51 AlaGlnAlaAlaValAlaArgPheGlnGluAlaAlaAsnLysGlnLysG 67  
201 GGAACCTGACGAGATCTCGACGAAATATTCGTCAGCGCGCGGTCCAACT 250  
67 nGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyr 84  
251 CGAGGGCGCGGAGGAGCAGCAGCAGCGCTGTCTCGCAAAATGGGCTTC 300  
84 erArgAlaAspGluGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 100  
seq\_name: /cgn2\_6/ptodata/2/paa/PCTUS\_COMB.pep:PCT-US99-03268-115

seq\_documentation\_block:

Sequence 115, Application PC/TUS9903268  
GENERAL INFORMATION:  
APPLICANT: Corixa Corporation  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY  
TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS  
NUMBER OF SEQUENCES: 355  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds, LLP

```
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US99/03268
; FILING DATE: 17-FEB-1999
; CLASSIFICATION: 406
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/025,197
; FILING DATE: 18-FEB-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura, A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 9532-0013-228
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-493-4935
; TELEFAX: 650-493-5556
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 115:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 100 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
;
PCT-US99-03268-115

alignment_scores:
  Quality: 486.00      Length: 100
  Ratio: 4.860        Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 99.000

alignment_block:
US-09-462-480-4 x PCT-US99-03268-115  ..
Align seg 1/1 to: PCT-US99-03268-115 from: 1 to: 100

1 ATGCAGAGATGAAGACCGATCGCGCTACCTCGCGGAGGAGGAGGTAA 50
|||||
1 MetAlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAs 17
51 TTTCGAGCGGATCTCCGCGACCTGAAACCCAGATCGACCGTGGAGT 100
|||||
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGluS 34
101 CGAGCGGAGGTTCGTCAGGCGCAGTGGCGCGCGCGCGGCGGACGCCC 150
|||||
34 erThrAlaGlySerLeuGlnGlyGlnIrpArgGlyAlaAlaGlyThrAla 50
151 GCCCAGGCGCGGTGGTGGCTTCCAAAGACGACCAATAGCAGAAGCA 200
|||||
51 AlaGlnAlaAlaValValArgPheGlnGluAlaAlaAsnLysGlnLysG 67
201 GGAACCTCGACGAGATCTCGACGAAATATTTCGTACGGCGCGCTCCAAT 250
|||||
67 nGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyrS 84
251 CGAGGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 300
|||||
84 erArgAlaAspGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnG 100
|||||

seq_name: /cgn_2/ptodata/2/paa/us087_COMB_pep:us-08-729-622-110
seq_documentation_block:
; Sequence 110, Application US/08729622
; GENERAL INFORMATION:
```

```
|||||
84 erArgAlaAspGluGlnGlnAlaLeuSerSerGlnMetGlyPhe 100
seq_name: /cgn2_5/ptodata/2/paa/US087_COMB.pep:US-08-730-510-115
seq_documentation_block:
; Sequence 115, Application US/08730510
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 137
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/730,510
; FILING DATE: 27-AUG-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 115:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 100 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-730-510-115

alignment_scores:
  Quality: 486.00      Length: 100
  Ratio: 4.860        Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.000

alignment_block:
US-09-462-480-4 x US-08-730-510-115 ..
Align seg 1/1 to: US-08-730-510-115 from: 1 to: 100

1 ATGGCAGAGATGAGACCGGATCGCGCTACCTCGGCGAGGAGCGGTAA 50
|||||
1 MetAlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAs 17
|||||
51 TTTCGAGCGGATCTCCGGCGACCTGAAACCCAGATCGACCGAGTGGAGT 100
|||||
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGluS 34
|||||
101 CGAGCGCAGGTTCTGTCAGGGCCAGTGGCGCGCGCGCGGCGGACGGCC 150
|||||
34 erThrAlaGlySerLeuGlnGlnGlnTrpArgGlyAlaAlaGlyThrAla 50
|||||
151 GCCAGGCGCGGTGGTGGCTTCCAGAGAGACGCCCAATAGCAGAGCA 200
|||||
```

```
|||||
51 AlaGlnAlaAlaValValArgPheGlnGlnAlaAlaAsnLysGlnLysG 67
|||||
201 GGAACCTCACGAGATCTCGACGAATATTCTGTCAGGCGCGGTCCCAACT 250
|||||
67 nGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyrS 84
|||||
251 CGAGGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGT 300
|||||
84 erArgAlaAspGluGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 100
|||||
seq_name: /cgn2_5/ptodata/2/paa/US089_COMB.pep:US-08-942-341-110
seq_documentation_block:
; Sequence 110, Application US/08942341
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; NUMBER OF SEQUENCES: 209
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/942,341
; FILING DATE: 01-OCT-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C7
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 110:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 100 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-942-341-110

alignment_scores:
  Quality: 486.00      Length: 100
  Ratio: 4.860        Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.000

alignment_block:
US-09-462-480-4 x US-08-942-341-110 ..
Align seg 1/1 to: US-08-942-341-110 from: 1 to: 100

1 ATGGCAGAGATGAGACCGGATCGCGCTACCTCGGCGAGGAGCGGTAA 50
|||||
1 MetAlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAs 17
|||||
51 TTTCGAGCGGATCTCCGGCGACCTGAAACCCAGATCGACCGAGTGGAGT 100
|||||
```

17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGluS 34  
101 CGAGCGCAGGTTCTTCAGCGCCAGTGGCGCGCGCGGGGACGCGCC 150  
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50  
151 GCCAGCGCGGTGGTGGCGCTTCCAAAGAACGACCAATAAGCAAGCA 200  
51 AlaGlnAlaAlaValAlaValArgPheGlnGlnAlaAlaAsnLysGlnLysG 67  
201 GGAAGTCTGACGAGATCTCGAGCAATATTCGTACGGCGCGGCTCCAAATGGCTTC 250  
67 nGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyrS 84  
251 CGAGCGCGCAGGAGCAGCAGCAGCGCTGCTCCGCAATGGCTTC 300  
84 erArgAlaAspGluGlnGlnGlnAlaLeuSerGlnMetGlyPhe 100  
seq\_name: /cgn2\_6/ptodata/2/paa/US089\_COMB pep:US-08-942-578-115

seq\_documentation\_block:  
; Sequence 115, Application US/08942578  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Campos-Neto, Antonio  
; APPLICANT: Houghton, Raymond  
; APPLICANT: Vedvick, Thomas S.  
; APPLICANT: Twardzik, Daniel R.  
; APPLICANT: Lodes, Michael J.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY  
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS  
; NUMBER OF SEQUENCES: 214  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED AND BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/942,578  
; FILING DATE: 01-OCT-1997  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.411C7  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 115:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 100 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
US-08-942-578-115

alignment\_scores:  
Quality: 486.00 Length: 100  
Ratio: 4.860 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 99.000  
alignment\_block:

US-09-462-480-4 x US-08-942-578-115  
Align seg 1/1 to: US-08-942-578-115 from: 1 to: 100  
1 ATGCGCAGAGATGAAGACCGATGCGCTACCTCGGGCAGGAGGAGGTAA 50  
1 MetAlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAs 17  
51 TTTCTGACGCGGATCTCCGCGCACCTGAAACCCAGATCGACGAGTGGAGT 100  
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGluS 34  
101 CGAGCGCAGGTTCTTCAGCGCCAGTGGCGCGCGGGGACGCGCC 150  
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50  
151 GCCAGCGCGGTGGTGGCGCTTCCAAAGAACGACCAATAAGCAAGCA 200  
51 AlaGlnAlaAlaValAlaValArgPheGlnGlnAlaAlaAsnLysGlnLysG 67  
201 GGAAGTCTGACGAGATCTCGAGCAATATTCGTACGGCGCGGCTCCAAAT 250  
67 nGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyrS 84  
251 CGAGCGCGCAGGAGCAGCAGCAGCGCTGCTCCGCAATGGCTTC 300  
84 erArgAlaAspGluGlnGlnGlnAlaLeuSerGlnMetGlyPhe 100  
seq\_name: /cgn2\_6/ptodata/2/paa/US090\_COMB pep:US-09-024-753-110

seq\_documentation\_block:  
; Sequence 110, Application US/09024753  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Campos-Neto, Antonio  
; APPLICANT: Houghton, Raymond  
; APPLICANT: Vedvick, Thomas S.  
; APPLICANT: Twardzik, Daniel R.  
; APPLICANT: Lodes, Michael J.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF  
; NUMBER OF SEQUENCES: 236  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED AND BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/024,753  
; FILING DATE: 18-FEB-1998  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.417C8  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 110:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 100 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
US-09-024-753-110

```
alignment_scores:
  Quality: 486.00      Length: 100
  Ratio: 4.860        Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.000

alignment_block:
US-09-462-480-4 x US-09-024-753-110 ..
Align seg 1/1 to: US-09-024-753-110 from: 1 to: 100

1 ATGCGAGAGATGAAGACCGATGCCCTACCTCGCGCAGGAGGAGTAA 50
1 MetAlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAs 17
51 TTTCGAGCGGATCTCCGGCGACCTGAAACCCAGATCCAGCAGGTGGAGT 100
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGluS 34
101 CGACGCGAGGTTCGTTCGAGGCGCAGTGGCGCGCGCGGAGCGGCC 150
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50
151 GCCCGAGCGCGGTGGTGGCTTCCAGAACGACGCAATTAAGCAGAACCA 200
51 AlaGlnAlaAlaValArgPheGlnGluAlaAlaAsnLysGlnLysG 67
201 GGAACCTCGACGAGATCTCGACGAATATTCGTTCAGGCGCGCTCCCAAT 250
67 nGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyrS 84
251 CGAGGCGCGCAGGAGCAGCAGGCGCGCTGTCTCCCAATGGGCTTC 300
84 erArgAlaAspGluGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 100
seq_name: /cgn2_6/ptodata/2/paa/US090_COMB.pap:US-09-025-197-115

seq_documentation_block:
; Sequence 115, Application US/09025197
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedrick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/025,197
; FILING DATE: 18-FEB-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C8
; TELECOMMUNICATION INFORMATION:
```

```
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 115:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 100 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-09-025-197-115

alignment_scores:
  Quality: 486.00      Length: 100
  Ratio: 4.860        Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.000

alignment_block:
US-09-462-480-4 x US-09-025-197-115 ..
Align seg 1/1 to: US-09-025-197-115 from: 1 to: 100

1 ATGCGAGAGATGAAGACCGATGCCCTACCTCGCGCAGGAGGAGTAA 50
1 MetAlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAs 17
51 TTTCGAGCGGATCTCCGGCGACCTGAAACCCAGATCCAGCAGGTGGAGT 100
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGluS 34
101 CGACGCGAGGTTCGTTCGAGGCGCAGTGGCGCGCGCGGAGCGGCC 150
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50
151 GCCCGAGCGCGGTGGTGGCTTCCAGAACGACGCAATTAAGCAGAACCA 200
51 AlaGlnAlaAlaValArgPheGlnGluAlaAlaAsnLysGlnLysG 67
201 GGAACCTCGACGAGATCTCGACGAATATTCGTTCAGGCGCGCTCCCAAT 250
67 nGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyrS 84
251 CGAGGCGCGCAGGAGCAGCAGGCGCGCTGTCTCCCAATGGGCTTC 300
84 erArgAlaAspGluGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 100
seq_name: /cgn2_6/ptodata/2/paa/US090_COMB.pap:US-09-072-596-110

seq_documentation_block:
; Sequence 110, Application US/09072596
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedrick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 110:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 100 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-09-072-596-110

alignment_scores:
    Quality: 486.00      Length: 100
    Ratio: 4.860        Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 99.000

alignment_block:
US-09-462-480-4 x US-09-072-596-110 ..
Align seg 1/1 to: US-09-072-596-110 from: 1 to: 100

1 ATGGCAGAGATGAAGACCGATCGCGCTACCTCGGCGGAGGAGGAGGTAA 50
1 MetAlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAs 17
51 TTTCGAGCGGATCTCCGCGGACCTGAAACCAGATCGACCGAGTGGAGT 100
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGlu 34
101 CGACGGCAGGTTCGTTGTCAGGCGCAGTGGCGGCGGCGGCGGCGGCC 150
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50
151 GCCCAGCGCGGCGGTGGTGCCTTCCAAGAAGCAGCCCAATAAGCAGAACA 200
51 AlaGlnAlaAlaValAlaArgPheGlnGlnAlaAlaAsnLysGlnLysG 67
201 GGAAGTCAGCAGATCTCGACGAATATTCGTACGCGCGGCGGCTCAATACT 250
67 nGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyr 84
251 CGAGGCGCGCAGGAGCAGCAGGCGGCTGCTCCTCGCAATGGGCTTC 300
84 erArgAlaAspGluGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 100

seq_name: /cgn2_6/ptodata/2/paa/US090_COMB.pap:US-09-072-967-115

seq_documentation_block:
; Sequence 115, Application US/09072967
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP

```

```

; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 115:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 100 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-09-072-967-115

alignment_scores:
    Quality: 486.00      Length: 100
    Ratio: 4.860        Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 99.000

alignment_block:
US-09-462-480-4 x US-09-072-967-115 ..
Align seg 1/1 to: US-09-072-967-115 from: 1 to: 100

1 ATGGCAGAGATGAAGACCGATCGCGCTACCTCGGCGGAGGAGGAGGTAA 50
1 MetAlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAs 17
51 TTTCGAGCGGATCTCCGCGGACCTGAAACCAGATCGACCGAGTGGAGT 100
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGlu 34
101 CGACGGCAGGTTCGTTGTCAGGCGCAGTGGCGGCGGCGGCGGCGGCC 150
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50
151 GCCCAGCGCGGCGGTGGTGCCTTCCAAGAAGCAGCCCAATAAGCAGAACA 200
51 AlaGlnAlaAlaValAlaArgPheGlnGlnAlaAlaAsnLysGlnLysG 67
201 GGAAGTCAGCAGATCTCGACGAATATTCGTACGCGCGGCGGCTCAATACT 250
67 nGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyr 84
251 CGAGGCGCGCAGGAGCAGCAGGCGGCTGCTCCTCGCAATGGGCTTC 300
84 erArgAlaAspGluGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 100

seq_name: /cgn2_6/ptodata/2/paa/US096_COMB.pap:US-09-615-947-1

seq_documentation_block:
; Sequence 1, Application US/09615947
; GENERAL INFORMATION:
; APPLICANT: ANDERSEN, Peter et al.
; TITLE OF INVENTION: Tuberculosis vaccine and diagnostics
; TITLE OF INVENTION: based on the Mycobacterium tuberculosis esat-6 gene family
; FILE REFERENCE: 0459-0462P

```

```
; CURRENT APPLICATION NUMBER: US/09/615,947
; CURRENT FILING DATE: 2001-06-04
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: fastseq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 100
; TYPE: PRT
; ORGANISM: M.Tuberculosis
US-09-615-947-1

alignment_scores:
  Quality: 486.00      Length: 100
  Ratio: 4.860         Gaps: 0
  Percent Similarity: 100.000      Percent Identity: 99.000

alignment_block:
US-09-462-480-4 x US-09-615-947-1
..
Align seg 1/1 to: US-09-615-947-1 from: 1 to: 100

1 ATGCAGAGATGAAGACCGATCCGCTACCTCGGCGAGGCGAGGTAA 50
|||||
1 MetAlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAs 17
|||||
51 TTTCGAGCGGATCTCCGGCGACTGAAACCCAGATCGACCGAGTGGAGT 100
|||||
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGlu 34
|||||
101 CGACGCGAGTTCGTTCGAGGCGCCAGTGGCGGGCGCGGGGAGCGCC 150
|||||
34 erThAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50
|||||
151 GCCAGCGCGCGGTGTCGCGCTTCCAAGACGAGCCCAATAGCAAGCA 200
|||||
51 AlaGlnAlaAlaValValArgPheGlnGluAlaAlaAsnLysGlnLysG 67
|||||
201 GGAACCTGACGAGATCTCGACGAATATTGTCAGCGCGCGTCCCAATACT 250
|||||
67 nGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyr 84
|||||
251 CGAGGCGCGAGGAGGAGGAGGAGGCGCTGCTCGCAATGGGCTTC 300
|||||
84 erArgAlaAspGluGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 100
|||||

seq_name: /cgn2_6/ptodata/2/paa/US097_COMB.pep:US-09-724-685-115

seq_documentation_block:
; Sequence 115, Application US/09724685
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; Skeiky, Yasir
; Dillon, Davin C.
; Campos-Neto, Antonio
; TITLE OF INVENTION: Compounds and Methods for
; Immunotherapy and Diagnosis of Tuberculosis
; NUMBER OF SEQUENCES: 155
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/724,685
; FILING DATE: 28-Nov-2000
; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/523,436
; FILING DATE: 01-SEP-1995
; APPLICATION NUMBER: US 08/533,634
; FILING DATE: 22-SEP-1995
; APPLICATION NUMBER: US 08/620,874
; FILING DATE: 22-MAR-1996
; APPLICATION NUMBER: US 08/659,683
; FILING DATE: 05-JUN-1996
; APPLICATION NUMBER: US 08/680,574
; FILING DATE: 12-JUL-1996
; APPLICATION NUMBER: WO PCT/US96/14674
; FILING DATE: 30-AUG-1996
; APPLICATION NUMBER: US 08/730,511
; FILING DATE: 11-OCT-1996
; APPLICATION NUMBER: US 08/818,112
; FILING DATE: 13-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Bastian, Kevin L.
; REGISTRATION NUMBER: 34,774
; REFERENCE/DOCKET NUMBER: 014058-008561US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 115:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 100 amino acids
; TYPE: amino acid
; STRANDEDNESS: <unknown>
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 115:
US-09-724-685-115

alignment_scores:
  Quality: 486.00      Length: 100
  Ratio: 4.860         Gaps: 0
  Percent Similarity: 100.000      Percent Identity: 99.000

alignment_block:
US-09-462-480-4 x US-09-724-685-115
..
Align seg 1/1 to: US-09-724-685-115 from: 1 to: 100

1 ATGCAGAGATGAAGACCGATCCGCTACCTCGGCGAGGCGAGGTAA 50
|||||
1 MetAlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAs 17
|||||
51 TTTCGAGCGGATCTCCGGCGACTGAAACCCAGATCGACCGAGTGGAGT 100
|||||
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGlu 34
|||||
101 CGACGCGAGTTCGTTCGAGGCGCCAGTGGCGGGCGCGGGGAGCGCC 150
|||||
34 erThAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50
|||||
151 GCCAGCGCGCGGTGTCGCGCTTCCAAGACGAGCCCAATAGCAAGCA 200
|||||
51 AlaGlnAlaAlaValValArgPheGlnGluAlaAlaAsnLysGlnLysG 67
|||||
201 GGAACCTGACGAGATCTCGACGAATATTGTCAGCGCGCGTCCCAATACT 250
|||||
67 nGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyr 84
|||||
251 CGAGGCGCGAGGAGGAGGAGGCGCTGCTCGCAATGGGCTTC 300
|||||
84 erArgAlaAspGluGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 100
|||||

seq_name: /cgn2_6/ptodata/2/paa/US100_COMB.pep:US-10-084-843-115

seq_documentation_block:
; Sequence 115, Application US/10084843
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; Skeiky, Yasir
; Dillon, Davin C.
; Campos-Neto, Antonio
; TITLE OF INVENTION: Compounds and Methods for
; Immunotherapy and Diagnosis of Tuberculosis
; NUMBER OF SEQUENCES: 155
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/724,685
; FILING DATE: 28-Nov-2000
; CLASSIFICATION: <Unknown>
```



APPLICANT: Reed, Steven G.  
Skeiky, Yasir A.W.  
Dillon, Davin C.  
Campos-Neto, Antonio  
Houghton, Raymond  
Vedvick, Thomas S.  
Twardzik, Daniel R.  
Lodes, Michael J.  
Hendrickson, Ronald C.  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY  
AND DIAGNOSIS OF TUBERCULOSIS  
NUMBER OF SEQUENCES: 355  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/084,843  
FILING DATE: 25-Feb-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/072,967  
FILING DATE: 05-MAY-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Makl, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.411C9  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 115:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 100 amino acids  
TYPE: amino acid  
STRANDEDNESS: <Unknown>  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 115:  
US-10-084-843-115

alignment\_scores:  
Quality: 486.00 Length: 100  
Ratio: 4.860 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 99.000  
alignment\_block:  
US-09-462-480-4 x US-10-084-843-115  
Align seq 1/1 to: US-10-084-843-115 from: 1 to: 100  
1 ATGCGAGATGAGACCGATCGCGTACCTCGGCGGAGGCGAGGTAA 50  
1 MetAlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAs 17  
51 TTTCGAGCGGATCTCCGCGACCTGAAACCCAGATCGACCGAGGTGGAGT 100  
17 nPheGluArgLysSerGlyAspLeuLysThrGlnIleAspGlnValGluS 34  
101 CGACGCGAGGTTCGTTCAGGGCCAGTGGCGCGCGCGGCGGACGCGCC 150  
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50  
151 GCCCAGCGCGCGGTGGTGGCGTTCACAGACGACGCCAATAGCAGAGCA 200  
51 AlaGlnAlaAlaValValArgPheGlnGlnGluAlaAlaAsnLysGlnLysG 67

201 GGAAGCTCGAGGATCTCGACGAATATTTCGTACGGCGGCGTCCAATACT 250  
|||||  
67 nGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyrS 84  
251 CGAGGGCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGTTC 300  
|||||  
84 erArgAlaAspGluGluGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 100



OM of: US-09-462-480-4 to: Pending\_Patents\_AA\_New:\* out\_format : pfs  
Date: Jul 22, 2002 1:37 AM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:  
-MODEL=frame+n2p.spool/US09462480/runat\_18072002\_164419\_19498/app\_query.fasta\_1.2850  
-DB=Pending\_Patents\_AA\_New -QFMT=fastan -SUFFIX=rapn  
-GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000  
-LOOPEXT=0.000 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000  
-XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000  
-YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1  
-MATRIX=blosom62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200  
-THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pfs -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09462480\_RCGN1\_1\_93 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

Search information block:  
Query: US-09-462-480-4  
Query length: 302  
Database: Pending\_Patents\_AA\_New:\*  
Database sequences: 378952  
Database length: 124292526  
Search time (sec): 104.950000

score_list:	Sequence	Strd Orig	zScore	EScore	Len	Documentation
/cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-116-492A-5 +	492.00	756.62	6.1e-34	10		
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-140-045-5 +	492.00	756.62	6.1e-34	100		
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-080-170B-639 +	486.00	747.47	2.0e-33	49		
/cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-116-492A-6 +	247.00	389.04	3.7e-13	49		
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-140-045-6 +	247.00	389.04	3.7e-13	49		
/cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-116-492A-7 +	205.00	326.30	1.4e-09	42		
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-140-045-7 +	205.00	326.30	1.4e-09	42		
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-09-116-492A-28 +	187.00	291.40	5.0e-08	10		
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-140-045-28 +	187.00	291.40	5.0e-08	10		
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-080-170B-12 +	187.00	291.40	5.0e-08	10		
/cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-116-492A-12 +	137.00	226.05	0.0008	2		
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-140-045-12 +	137.00	226.05	0.0008	28		
/cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-935-625-17077 -	119.00	169.55	0.0369	10		
/cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-935-625-17076 -	119.00	169.35	0.0370	10		
/cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-935-625-17480 -	119.00	169.23	0.0371	10		
/cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-935-625-17479 -	119.00	169.03	0.0372	10		
/cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-935-625-17075 -	119.00	168.91	0.0372	10		
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-155-881-71478 -	119.00	168.80	0.0373	10		
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-155-881-7165 +	109.50	167.32	0.2052	10		
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-155-881-26387 +	102.50	146.90	0.9045	10		
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-155-881-26077 +	100.50	148.07	1.27	10		
/cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-116-492A-8 +	100.00	172.50	1.05	20		
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-140-045-8 +	100.00	172.50	1.05	20		
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-155-881-10947 +	99.00	154.59	1.54	10		
/cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep:US-60-389-987-2231 +	98.50	129.65	2.25	10		
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-155-881-10418 -	96.50	147.24	2.62	10		
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-155-881-9649 -	96.00	149.52	2.79	10		
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-155-881-9768 +	96.00	144.33	2.96	10		
/cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep:US-60-389-987-2448 -	96.00	132.37	3.40	10		
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-155-881-11792 +	95.50	140.65	3.38	10		
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-155-881-11791 -	95.50	140.65	3.38	10		
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-155-881-7219 -	95.00	141.45	3.66	10		
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-155-881-10185 -	95.00	137.68	3.82	10		
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-155-881-7135 -	94.50	149.11	3.66	10		
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-155-881-10473 -	94.50	142.69	3.94	10		
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-155-881-27136 -	93.50	141.94	4.75	10		
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-155-881-17173 +	93.00	140.95	5.25	10		
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-155-881-12231 -	93.00	139.43	5.34	10		
/cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep:PCT-US02-06388-2 -	92.50	140.49	5.77	220		
/cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep:US-60-389-987-2409 -	92.50	140.11	5.79	10		

seq\_name: /cgn2\_6/ptodata/2/paa/US10\_NEW\_COMB.pep:US-09-116-492A-5  
seq\_documentation\_block:  
; Sequence 5, Application US/09116492A  
; GENERAL INFORMATION:  
; APPLICANT: GICQUEL, BRIGITTE  
; APPLICANT: BERTHET, FRANCOIS-XAVIER  
; APPLICANT: ANDERSEN, PETER  
; APPLICANT: RASMUSSEN, PETER B  
; TITLE OF INVENTION: POLYNUCLEOTIDE FUNCTIONALLY CODING FOR THE LHP PROTEIN FROM M  
; TITLE OF INVENTION: TUBERCULOSIS, ITS BIOLOGICALLY ACTIVE DERIVATIVE FRAGMENTS, I  
; FILE REFERENCE: 0660-0137-27X  
; CURRENT APPLICATION NUMBER: US/09/116,492A  
; CURRENT FILING DATE: 1998-07-16  
; PRIOR APPLICATION NUMBER: 60/252,631  
; PRIOR FILING DATE: 1997-07-16  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 5  
; LENGTH: 100  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis  
US-09-116-492A-5

```

alignment_scores:
  Quality: 492.00      Length: 100
  Ratio: 4.920        Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
  US-09-462-480-4 x US-09-116-492A-5  ..

  Align seg 1/1 to: US-09-116-492A-5 from: 1 to: 100

1 ATGCAGAGATGAAGACCGATGCGGCTACCTCGGCGAGGAGCAGGTAA 50
|||||
1 MetAlaGluMetLysThrAspAlaAlaThrLeuGlyGlnGluAlaGlyAs 17

51 TTTTCAGCGGATCTCCGGCGACCTGAAACCCAGATCGACCGAGTGGAGT 100
|||||
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGlu 34

101 CGACGCGAGGTTCTTCAGGCGCGAGTGGCGCGCGCGGCGGAGCGCC 150
|||||
34 erThrAlaGlySerLeuGlnGlyGlnIleArgGlyAlaAlaGlyThrAla 50

151 GCCAGCGCGCGGTTGGTGGCTTCCAAAGAACGACGCCAATAAGCAAGCA 200
|||||
51 AlaGlnAlaAlaValAlaArgPheGlnGluAlaAlaAsnLysGlnLysG 67

201 GGAACCTGACGAGATCTCGAGCATATTCGTAGCGCGGCTCCAACTACT 250
|||||
67 nGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyr 84

251 CGAGGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 300
|||||
84 erArgAlaAspGluGlnGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 100

seq_name: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-140-045-5

seq_documentation_block:
; Sequence 5, Application US/10140045
; GENERAL INFORMATION:
; APPLICANT: GICQUEL, BRIGITTE
; APPLICANT: BERTHET, FRANCOIS-XAVIER

```

```
; APPLICANT: ANDERSEN, PETER
; APPLICANT: RASMUSSEN, PETER B
; TITLE OF INVENTION: POLYNUCLEOTIDE FUNCTIONALLY CODING FOR THE LHP PROTEIN FROM MYCOBACTERIUM TUBERCULOSIS, ITS BIOLOGICALLY ACTIVE DERIVATIVE FRAGMENTS, AS WELL AS THE USE OF SUCH COMPOSITIONS IN THE TREATMENT OF TUBERCULOSIS
; TITLE OF INVENTION: TUBERCULOSIS, ITS BIOLOGICALLY ACTIVE DERIVATIVE FRAGMENTS, AS WELL AS THE USE OF SUCH COMPOSITIONS IN THE TREATMENT OF TUBERCULOSIS
; FILE REFERENCE: 0660-0137-27X
; CURRENT APPLICATION NUMBER: US/10/140,045
; CURRENT FILING DATE: 2002-05-08
; PRIOR APPLICATION NUMBER: US/09/116,492A
; PRIOR FILING DATE: 1998-07-16
; PRIOR APPLICATION NUMBER: 60/252,631
; PRIOR FILING DATE: 1997-07-16
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; US-10-140-045-5

alignment_scores:
  Quality: 492.00      Length: 100
  Ratio: 4.920        Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-462-480-4 x US-10-140-045-5 ..
Align seg 1/1 to: US-10-140-045-5 from: 1 to: 100

1 ATGCGAGAGATGAAGACCGATGCCCTACCTCGGCGAGGAGGAGTAA 50
1 MetAlaGluMetLysThrAspAlaAlaThrLeuGlyGlnGluAlaGlyAs 17
51 TTTCGAGCGGATCTCCGGCGACCTGAAACCCAGATCGACGAGTGGAGT 100
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGlu 34
101 CGAGCGCAGGTTGTTGCGGCGAGTGGCGGCGGCGGCGGCGGCGGCG 150
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50
151 GCCGAGCGCGGTTGCGGCTTCCAAAGCAGCCCAATAGCAGCAAGCA 200
51 AlaGlnAlaAlaValArgPheGlnGlnAlaAlaAsnLysGlnLysG 67
201 GGAACGACGAGATCTCGACGAATATTCGTACGCGGCGGCGGCGGCG 250
67 nGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyr 84
151 GCCGAGCGCGGTTGCGGCTTCCAAAGCAGCCCAATAGCAGCAAGCA 200
51 AlaGlnAlaAlaValArgPheGlnGlnAlaAlaAsnLysGlnLysG 67
201 GGAACGACGAGATCTCGACGAATATTCGTACGCGGCGGCGGCGGCG 250
67 nGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyr 84
251 CGAGCGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 300
84 erArgAlaAspGluGlnGlnGlnGlnAlaLeuSerGlnMetGlyPhe 100

seq_name: /cgn2_6/ptodata/2/paa/us10_NEW_COMB.pap:us-10-080-170B-639

seq_documentation_block:
; Sequence 639, Application US/10080170B
; GENERAL INFORMATION:
; APPLICANT: COLE, S.T.
; TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR TREATMENT OF MYCOBACTERIOSES
; FILE REFERENCE: 03495.0218
; CURRENT APPLICATION NUMBER: US/10/080,170B
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: 60/270,123
; PRIOR FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 652
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 639
; LENGTH: 100
```

```
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; US-10-080-170B-639

alignment_scores:
  Quality: 486.00      Length: 100
  Ratio: 4.860        Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 99.000

alignment_block:
US-09-462-480-4 x US-10-080-170B-639 ..
Align seg 1/1 to: US-10-080-170B-639 from: 1 to: 100

1 ATGCGAGAGATGAAGACCGATGCCCTACCTCGGCGAGGAGGAGTAA 50
1 MetAlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAs 17
51 TTTCGAGCGGATCTCCGGCGACCTGAAACCCAGATCGACGAGTGGAGT 100
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGlu 34
101 CGAGCGCAGGTTGTTGCGGCGAGTGGCGGCGGCGGCGGCGGCGGCG 150
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50
151 GCCGAGCGCGGTTGCGGCTTCCAAAGCAGCCCAATAGCAGCAAGCA 200
51 AlaGlnAlaAlaValArgPheGlnGlnAlaAlaAsnLysGlnLysG 67
201 GGAACGACGAGATCTCGACGAATATTCGTACGCGGCGGCGGCGGCG 250
67 nGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyr 84
251 CGAGCGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 300
84 erArgAlaAspGluGlnGlnGlnGlnAlaLeuSerGlnMetGlyPhe 100

seq_name: /cgn2_6/ptodata/2/paa/us09_NEW_COMB.pap:us-09-116-492A-6

seq_documentation_block:
; Sequence 6, Application US/09116492A
; GENERAL INFORMATION:
; APPLICANT: GICQUEL, BRIGITTE
; APPLICANT: BERTHET, FRANCOIS-XAVIER
; APPLICANT: ANDERSEN, PETER
; APPLICANT: RASMUSSEN, PETER B
; TITLE OF INVENTION: POLYNUCLEOTIDE FUNCTIONALLY CODING FOR THE LHP PROTEIN FROM MYCOBACTERIUM TUBERCULOSIS, ITS BIOLOGICALLY ACTIVE DERIVATIVE FRAGMENTS, AS WELL AS THE USE OF SUCH COMPOSITIONS IN THE TREATMENT OF TUBERCULOSIS
; FILE REFERENCE: 0660-0137-27X
; CURRENT APPLICATION NUMBER: US/09/116,492A
; CURRENT FILING DATE: 1998-07-16
; PRIOR APPLICATION NUMBER: 60/252,631
; PRIOR FILING DATE: 1997-07-16
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 49
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; US-09-116-492A-6

alignment_scores:
  Quality: 247.00      Length: 49
  Ratio: 5.041        Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-462-480-4 x US-09-116-492A-6 ..
```

Align seg 1/1 to: US-09-116-492A-6 from: 1 to: 49

1 ATGCGCAGATGAGACCGATGCGCTACCTCGGCGAGGCGAGTAA 50  
|||||  
1 MetAlaGluMetLysThrAspAlaAlaThrLeuGlyGlnGluAlaGlyAs 17  
|||||  
51 TTTTCGACGGATCTCCGGCGACCTGAAACCCAGATCGACCGAGTGGAGT 100  
|||||  
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGluS 34  
|||||  
101 CGACGGCAGGTTCGTTCCAGGGCAGTGGCGCGCGCGGGGAGC 147  
|||||  
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThr 49  
|||||

seq\_name: /cgn2\_6/ptodata/2/paa/US10\_NEW\_COMB.pep:US-10-140-045-6

seq\_documentation\_block:  
; Sequence 6, Application US/10140045  
; GENERAL INFORMATION:  
; APPLICANT: GICQUEL, BRIGITTE  
; APPLICANT: BERTHET, FRANCOIS-XAVIER  
; APPLICANT: ANDERSEN, PETER  
; APPLICANT: RASMUSSEN, PETER B  
; TITLE OF INVENTION: POLYNUCLEOTIDE FUNCTIONALLY CODING FOR THE LHP PROTEIN FROM MYCOBACTERIUM TUBERCULOSIS, ITS BIOLOGICALLY ACTIVE DERIVATIVE FRAGMENTS, AS WELL AS  
; TITLE OF INVENTION: USING THE SAME  
; FILE REFERENCE: 0660-0137-27X  
; CURRENT APPLICATION NUMBER: US/10/140,045  
; PRIOR FILING DATE: 2002-05-08  
; PRIOR FILING DATE: 1998-07-16  
; PRIOR FILING DATE: 1998-07-16  
; PRIOR FILING DATE: 1997-07-16  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 6  
; LENGTH: 49  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis  
US-10-140-045-6

alignment\_scores:  
Quality: 247.00 Length: 49  
Ratio: 5.041 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-462-480-4 x US-10-140-045-6  
Align seg 1/1 to: US-10-140-045-6 from: 1 to: 49

1 ATGCGCAGATGAGACCGATGCGCTACCTCGGCGAGGCGAGTAA 50  
|||||  
1 MetAlaGluMetLysThrAspAlaAlaThrLeuGlyGlnGluAlaGlyAs 17  
|||||  
51 TTTTCGACGGATCTCCGGCGACCTGAAACCCAGATCGACCGAGTGGAGT 100  
|||||  
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGluS 34  
|||||  
101 CGACGGCAGGTTCGTTCCAGGGCAGTGGCGCGCGGGGAGC 147  
|||||  
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThr 49  
|||||

seq\_name: /cgn2\_6/ptodata/2/paa/US09\_NEW\_COMB.pep:US-09-116-492A-7

seq\_documentation\_block:  
; Sequence 7, Application US/09116492A  
; GENERAL INFORMATION:  
; APPLICANT: GICQUEL, BRIGITTE  
; APPLICANT: BERTHET, FRANCOIS-XAVIER  
; APPLICANT: ANDERSEN, PETER  
; APPLICANT: RASMUSSEN, PETER B

; TITLE OF INVENTION: POLYNUCLEOTIDE FUNCTIONALLY CODING FOR THE LHP PROTEIN FROM MYCOBACTERIUM TUBERCULOSIS, ITS BIOLOGICALLY ACTIVE DERIVATIVE FRAGMENTS, AS WELL AS  
; TITLE OF INVENTION: USING THE SAME  
; FILE REFERENCE: 0660-0137-27X  
; CURRENT APPLICATION NUMBER: US/09/116,492A  
; PRIOR FILING DATE: 1998-07-16  
; PRIOR FILING DATE: 1998-07-16  
; PRIOR FILING DATE: 1997-07-16  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 7  
; LENGTH: 42  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis  
US-09-116-492A-7

alignment\_scores:  
Quality: 205.00 Length: 42  
Ratio: 4.881 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-462-480-4 x US-09-116-492A-7  
Align seg 1/1 to: US-09-116-492A-7 from: 1 to: 42

175 CAAGAAGCAGCCCAATAGCAGGAGCAAGCACTCGACGAGATCTCGACGAA 224  
|||||  
1 GlnGluAlaAlaAsnLysGlnLysGlnGluLeuAspGluIleSerThrAs 17  
|||||  
225 TATTCGTGAGCGCGCGTCCCAATCTCGAGGCGCGAGGAGGAGCAGCAGC 274  
|||||  
17 nIleArgGlnAlaGlyValGlnTyrSerArgAlaAspGluGlnGlnG 34  
|||||  
275 AGCGCTGTCTCTCCGAAATGGGCTTC 300  
|||||  
34 nAlaLeuSerSerGlnMetGlyPhe 42  
|||||

seq\_name: /cgn2\_6/ptodata/2/paa/US10\_NEW\_COMB.pep:US-10-140-045-7

seq\_documentation\_block:  
; Sequence 7, Application US/10140045  
; GENERAL INFORMATION:  
; APPLICANT: GICQUEL, BRIGITTE  
; APPLICANT: BERTHET, FRANCOIS-XAVIER  
; APPLICANT: ANDERSEN, PETER  
; APPLICANT: RASMUSSEN, PETER B  
; TITLE OF INVENTION: POLYNUCLEOTIDE FUNCTIONALLY CODING FOR THE LHP PROTEIN FROM MYCOBACTERIUM TUBERCULOSIS, ITS BIOLOGICALLY ACTIVE DERIVATIVE FRAGMENTS, AS WELL AS  
; TITLE OF INVENTION: USING THE SAME  
; FILE REFERENCE: 0660-0137-27X  
; CURRENT APPLICATION NUMBER: US/10/140,045  
; CURRENT FILING DATE: 2002-05-08  
; PRIOR APPLICATION NUMBER: US/09/116,492A  
; PRIOR FILING DATE: 1998-07-16  
; PRIOR FILING DATE: 1998-07-16  
; PRIOR FILING DATE: 1997-07-16  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 7  
; LENGTH: 42  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis  
US-10-140-045-7

alignment\_scores:  
Quality: 205.00 Length: 42  
Ratio: 4.881 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

```
US-09-462-480-4 x US-10-140-045-7
Align seg 1/1 to: US-10-140-045-7 from: 1 to: 42

175 CAAGAAGCAGCAATAGCAGACAGCAAGATCTCGACGAGTCTCGACGAA 224
|||||
1 GlnGluAlaAlaAnLysGlnGlnGlnLeuLeuSerGlnIleSerThrAs 17
|||||
225 TATTCGTCAGCGCGGTCCCAATCTACTCGAGGGCGGACGAGGAGCAGCAGC 274
|||||
17 nIleArgGlnAlaGlyValGlnTrpSerArgAlaAspGluGlnGlnG 34
|||||
275 AGGCGCTGTCTCCCAATAGCGCTTC 300
|||||
34 nAlaLeuSerSerGlnMetGlyPhe 42

seq_name: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-116-492A-28

seq_documentation_block:
; Sequence 28, Application US/09116492A
; GENERAL INFORMATION:
; APPLICANT: GICOUEL, BRIGITTE
; APPLICANT: BERTHET, FRANCOIS-XAVIER
; APPLICANT: ANDERSEN, PETER
; APPLICANT: RASMUSSEN, PETER B
; TITLE OF INVENTION: POLYNUCLEOTIDE FUNCTIONALLY CODING FOR THE LHP PROTEIN FROM MYCOBACTERIUM LEPRAE, ITS BIOLOGICALLY ACTIVE DERIVATIVE FRAGMENTS, AS WELL AS A TOOL FOR IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR THERAPY OF TUBERCULOSIS
; TITLE OF INVENTION: USING THE SAME
; FILE REFERENCE: 0660-0137-27X
; CURRENT APPLICATION NUMBER: US/09/116,492A
; CURRENT FILING DATE: 1998-07-16
; PRIOR APPLICATION NUMBER: 60/252,631
; PRIOR FILING DATE: 1997-07-16
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 28
; LENGTH: 100
; TYPE: PRT
; ORGANISM: Mycobacterium leprae
US-09-116-492A-28
```

```
alignment_scores:
  Quality: 187.00      Length: 100
  Ratio: 2.367        Gaps: 0
  Percent Similarity: 79.000  Percent Identity: 40.000

alignment_block:
US-09-462-480-4 x US-09-116-492A-28

Align seg 1/1 to: US-09-116-492A-28 from: 1 to: 100

1 ATGCGCAGATGAGACCGGATCGCGCTACCTCGGCGAGGAGCGGTAA 50
|||||
1 MetAlaGluMetIleThrGluAlaAlaIleLeuThrGlnGlnAlaIaGl 17
|||||
51 TTTCGAGCGGATCTCCGCGCATCTGAAACCCAGATCGACCGGTGGAGT 100
|||||
17 nPheAspGlnIleAlaSerGlyLeuSerGlnGluArgAsnPheValAspS 34
|||||
101 CGACGCGCAGGTTCGTTCGACGGCCAGTGGCGCGCGCGGCGGACGCC 150
|||
34 erIleGlyGlnSerPheGlnAsnThrTrpGluGlyGlnAlaAlaSerAla 50
|||||
151 GCCCAGCGCGGTGGTGGTTCCTCCAGAGACGACGACCAATAGCAGCA 200
|||
51 AlalaLeuGlyAlaLeuGlyArgPheAspGluAlaMetGlnAspGlnIleAr 67
|||||
201 GGAATCGACGAGATCTCGACGAATATTCGTCAGCGCGCGCTCCAATACT 250
|||||
67 glnLeuGluSerIleValAspLysLeuAsnArgSerGlyGlyAsnTyrT 84
|||||
251 CGAGGGCGCAGCAGCAGCAGCAGCGGTGTCTCGCAATGGGCTTC 300
|||
34 erIleGlyGlnSerPheGlnAsnThrTrpGluGlyGlnAlaAlaSerAla 50
|||||
151 GCCCAGCGCGGTGGTGGTTCCTCCAGAGACGACGACCAATAGCAGCA 200
|||
51 AlalaLeuGlyAlaLeuGlyArgPheAspGluAlaMetGlnAspGlnIleAr 67
|||||
201 GGAATCGACGAGATCTCGACGAATATTCGTCAGCGCGCGCTCCAATACT 250
|||||
67 glnLeuGluSerIleValAspLysLeuAsnArgSerGlyGlyAsnTyrT 84
|||||
251 CGAGGGCGCAGCAGCAGCAGCAGCGGTGTCTCGCAATGGGCTTC 300
```

```
seq_name: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-140-045-28

seq_documentation_block:
; Sequence 28, Application US/10140045
; GENERAL INFORMATION:
; APPLICANT: GICOUEL, BRIGITTE
; APPLICANT: BERTHET, FRANCOIS-XAVIER
; APPLICANT: ANDERSEN, PETER
; APPLICANT: RASMUSSEN, PETER B
; TITLE OF INVENTION: POLYNUCLEOTIDE FUNCTIONALLY CODING FOR THE LHP PROTEIN FROM MYCOBACTERIUM LEPRAE, ITS BIOLOGICALLY ACTIVE DERIVATIVE FRAGMENTS, AS WELL AS A TOOL FOR IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR THERAPY OF TUBERCULOSIS
; TITLE OF INVENTION: USING THE SAME
; FILE REFERENCE: 0660-0137-27X
; CURRENT APPLICATION NUMBER: US/10/140,045
; CURRENT FILING DATE: 2002-05-08
; PRIOR APPLICATION NUMBER: US/09/116,492A
; PRIOR FILING DATE: 1998-07-16
; PRIOR APPLICATION NUMBER: 60/252,631
; PRIOR FILING DATE: 1997-07-16
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 28
; LENGTH: 100
; TYPE: PRT
; ORGANISM: Mycobacterium leprae
US-10-140-045-28

alignment_scores:
  Quality: 187.00      Length: 100
  Ratio: 2.367        Gaps: 0
  Percent Similarity: 79.000  Percent Identity: 40.000

alignment_block:
US-09-462-480-4 x US-10-140-045-28

Align seg 1/1 to: US-10-140-045-28 from: 1 to: 100

1 ATGCGCAGATGAGACCGGATCGCGCTACCTCGGCGAGGAGCGGTAA 50
|||||
1 MetAlaGluMetIleThrGluAlaAlaIleLeuThrGlnGlnAlaIaGl 17
|||||
51 TTTCGAGCGGATCTCCGCGCATCTGAAACCCAGATCGACCGGTGGAGT 100
|||||
17 nPheAspGlnIleAlaSerGlyLeuSerGlnGluArgAsnPheValAspS 34
|||||
101 CGACGCGCAGGTTCGTTCGACGGCCAGTGGCGCGCGCGGCGGACGCC 150
|||
34 erIleGlyGlnSerPheGlnAsnThrTrpGluGlyGlnAlaAlaSerAla 50
|||||
151 GCCCAGCGCGGTGGTGGTTCCTCCAGAGACGACGACCAATAGCAGCA 200
|||
51 AlalaLeuGlyAlaLeuGlyArgPheAspGluAlaMetGlnAspGlnIleAr 67
|||||
201 GGAATCGACGAGATCTCGACGAATATTCGTCAGCGCGCGCTCCAATACT 250
|||||
67 glnLeuGluSerIleValAspLysLeuAsnArgSerGlyGlyAsnTyrT 84
|||||
251 CGAGGGCGCAGCAGCAGCAGCAGCGGTGTCTCGCAATGGGCTTC 300
|||
84 hrLysThrAspAspGluAlaAsnGlnLeuLeuSerSerLysMetAsnPhe 100

seq_name: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-080-170B-12

seq_documentation_block:
; Sequence 12, Application US/10080170B
; GENERAL INFORMATION:
; APPLICANT: COLE, S.T.
; TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR THERAPY OF TUBERCULOSIS
```

```
; TITLE OF INVENTION: TREATMENT OF MYCOBACTERIOSES
; FILE REFERENCE: 03495.0218
; CURRENT APPLICATION NUMBER: US/10/080,170B
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: 60/270,123
; PRIOR FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 652
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 100
; TYPE: PRT
; ORGANISM: Mycobacterium leprae
US-10-080-170B-12
```

```
alignment_scores:
  Quality: 187.00      Length: 100
  Ratio: 2.367         Gaps: 0
Percent Similarity: 79.000 Percent Identity: 40.000
```

```
alignment_block:
US-09-462-480-4 x US-10-080-170B-12 ..
```

```
Align seg 1/1 to: US-10-080-170B-12 from: 1 to: 100
```

```
1 ATGCAGAGATGAAGACCGATGCCCTACCTCGGCGAGGACGAGTAA 50
|||||  |||:|||||  |||  |||:|||||:
1 MetAlaGluMetIleThrGluAlaAlaIleLeuThrGlnAlaAlaG 17
51 TTTCGACGGATCTCCGGCGACCTGAACCCAGATCGACAGTGAGT 100
|||||:|||||:  |||:  |||:  |||:  |||:  |||:  |||:
17 nPheAspGlnIleAlaSerGlyLeuSerGlnGluArgAsnPheValAsp 34
101 CGAGCGGAGGTTCCTTCAGGCGCAGTGCGCGCGGCGGAGCGGCC 150
||  ||:  |||:|||||:  |||:|||||  |||:|||||:
34 erlLeGlyGlnSerPheGlnAsnThrTrpGluGlyGlnAlaIleAla 50
151 GCCAGCGCGGCGTGCGCTTCCAAAGACGACGCCCAATAGCAGCAAGCA 200
|||  ||:|||||:  |||:|||||  |||  |||  |||  |||  |||
51 AlaLeuGlyAlaLeuGlyArgPheAspGluAlaMetGlnAspGlnIleAr 67
201 GGAACCTCGAGATCTCGAGGAATATTCGTACGGCGCGGCTCAATACT 250
||:|||||:|||||  |||:|||||:|||||  |||  |||:
67 gGlnLeuGluSerIleValAspLysLeuAsnArgSerGlyGlyAsnTyr 84
251 CGAGGCGCGAGGACGACGAGCGCTGTCTCGCAAAATGGGCTTC 300
||:|||||:|||||  |||:|||||  |||:|||||:|||||:|||||
84 hrLysThrAspAspGluAlaAsnGlnLeuLeuSerSerLysMetAsnPhe 100
```

```
seq_name: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pap:US-09-116-492A-12
```

```
seq_documentation_block:
```

```
; Sequence 12, Application US/09116492A
```

```
; GENERAL INFORMATION:
```

```
; APPLICANT: GICQUEL, BRIGITTE
```

```
; APPLICANT: BERTHET, FRANCOIS-XAVIER
```

```
; APPLICANT: ANDERSEN, PETER
```

```
; APPLICANT: RASMUSSEN, PETER B
```

```
; TITLE OF INVENTION: POLYNUCLEOTIDE FUNCTIONALLY CODING FOR THE LHP PROTEIN FROM MYCOB
```

```
; TITLE OF INVENTION: TUBERCULOSIS, ITS BIOLOGICALLY ACTIVE DERIVATIVE FRAGMENTS, AS W
```

```
; FILE REFERENCE: 0660-0137-27X
```

```
; CURRENT APPLICATION NUMBER: US/09/116,492A
```

```
; PRIOR FILING DATE: 1998-07-16
```

```
; PRIOR APPLICATION NUMBER: 60/252,631
```

```
; NUMBER OF SEQ ID NOS: 39
```

```
; SOFTWARE: PatentIn version 3.1
```

```
; SEQ ID NO 12
```

```
; LENGTH: 28
```

```
; TYPE: PRT
```

```
; ORGANISM: Mycobacterium tuberculosis
```

```
US-09-116-492A-12
```

```
alignment_scores:
  Quality: 137.00      Length: 28
  Ratio: 4.893         Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-462-480-4 x US-09-116-492A-12 ..
```

```
Align seg 1/1 to: US-09-116-492A-12 from: 1 to: 28
```

```
217 TCGACGAATATTCGTACGCGCGGCTCCATACTCGAGGCGCGCAGGAGA 266
|||||  |||:|||||  |||  |||:|||||:
1 SerThrAsnIleArgGlnAlaGlyValGlnTyrSerArgAlaAspGluG 17
267 GCAGCAGCAGCGCTGCTCGCAAAATGGGCTTC 300
|||||  |||:|||||  |||  |||:|||||:
17 uGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 28
```

```
seq_name: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pap:US-10-140-045-12
```

```
seq_documentation_block:
```

```
; Sequence 12, Application US/10140045
```

```
; GENERAL INFORMATION:
```

```
; APPLICANT: GICQUEL, BRIGITTE
```

```
; APPLICANT: BERTHET, FRANCOIS-XAVIER
```

```
; APPLICANT: ANDERSEN, PETER
```

```
; APPLICANT: RASMUSSEN, PETER B
```

```
; TITLE OF INVENTION: POLYNUCLEOTIDE FUNCTIONALLY CODING FOR THE LHP PROTEIN FROM MY
```

```
; TITLE OF INVENTION: TUBERCULOSIS, ITS BIOLOGICALLY ACTIVE DERIVATIVE FRAGMENTS, A
```

```
; FILE REFERENCE: 0660-0137-27X
```

```
; CURRENT APPLICATION NUMBER: US/10/140,045
```

```
; PRIOR FILING DATE: 2002-05-08
```

```
; PRIOR APPLICATION NUMBER: US/09/116,492A
```

```
; PRIOR FILING DATE: 1998-07-16
```

```
; PRIOR APPLICATION NUMBER: 60/252,631
```

```
; NUMBER OF SEQ ID NOS: 39
```

```
; SOFTWARE: PatentIn version 3.1
```

```
; SEQ ID NO 12
```

```
; LENGTH: 28
```

```
; TYPE: PRT
```

```
; ORGANISM: Mycobacterium tuberculosis
```

```
US-10-140-045-12
```

```
alignment_scores:
  Quality: 137.00      Length: 28
  Ratio: 4.893         Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
```

```
alignment_block:
US-09-462-480-4 x US-10-140-045-12 ..
```

```
Align seg 1/1 to: US-10-140-045-12 from: 1 to: 28
```

```
217 TCGACGAATATTCGTACGCGCGGCTCCATACTCGAGGCGCGCAGGAGA 266
|||||  |||:|||||  |||  |||:|||||:
1 SerThrAsnIleArgGlnAlaGlyValGlnTyrSerArgAlaAspGluG 17
267 GCAGCAGCAGCGCTGCTCGCAAAATGGGCTTC 300
|||||  |||:|||||  |||  |||:|||||:
17 uGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 28
```

```
seq_name: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pap:US-09-935-625-17077
```

```
seq_documentation_block:
```

```
; Sequence 17077, Application US/09935625
```

```
; GENERAL INFORMATION:
```

```
; APPLICANT: N. ALEXANDROV et al.
```

```
; TITLE OF INVENTION: POLYNUCLEOTIDES, POLYPEPTIDES, CELLS, AND METHODS THEREOF CAPA
```

```
; TITLE OF INVENTION: MODULATING VARIOUS RESPONSES
; FILE REFERENCE: 2750-1481P
; CURRENT APPLICATION NUMBER: US/09/935,625
; CURRENT FILING DATE: 2001-08-24
; NUMBER OF SEQ ID NOS: 33136
; SEQ ID NO 17077
; LENGTH: 827
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: peptide
; LOCATION: 1..827
; OTHER INFORMATION: Ceres Seq. ID no. 2708208
US-09-935-625-17077
```

```
alignment_scores:
  Quality: 119.00      Length: 111
  Ratio: 2.052        Gaps: 6
  Percent Similarity: 52.252  Percent Identity: 35.135
```

alignment\_block:

US-09-462-480-4/rev x US-09-935-625-17077 ..

Align seg 1/1 to: US-09-935-625-17077 from: 1 to: 827

```
301 AGAAGCCATTGCGAGGACAGCGCTGCTGCTCTCT...CGTCGCC 255
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
361 ArgSerProLeuAlaLysArgGlyArgSerAspSerProGlyArgSerPr 377
|||||:|||||:|||||:|||||:|||||:|||||:|||||
254 CTCGAGTATTGGACCGCGGCTGACGAATATTCTCGTGGAGA..... 215
|||||:|||||:|||||:|||||:|||||:|||||:|||||
377 oSerProValAlaArgLeuArgAspProThrGlyAlaArgLeuProSerP 394
|||||:|||||:|||||:|||||:|||||:|||||:|||||
214 .....TCTCGTCGAGTTCTCTGCTTCTGCTTATTGCTGCTCTTTGGA 173
|||||:|||||:|||||:|||||:|||||:|||||:|||||
394 roSerIleGluGlnArgLeuProSerProProValAlaGlnArgLeuPro 410
|||||:|||||:|||||:|||||:|||||:|||||:|||||
172 AGCGCACACCG...CGGCTGGCGCGCTGCCGCCGCCGCCGCCAC 126
|||||:|||||:|||||:|||||:|||||:|||||:|||||
411 SerProProProArgArgAlaGlyLeuProSerProPro..... 423
125 TGGCCCTGCAACGACCTGCGTGCACCTCCACCTGGTCTGGTGGTTT 76
|||||:|||||:|||||:|||||:|||||:|||||:|||||
424 ....ProAlaGlnArgLeuProSerProProProArgArgAlaGlyLeuP 439
|||||:|||||:|||||:|||||:|||||:|||||:|||||
75 CAGGTCCCGCGGAGA.....TCCGCTCGAAATTACCTG..... 44
|||||:|||||:|||||:|||||:|||||:|||||:|||||
439 roSerProMetArgIleGlyGlySerHisAlaAlaAsnHisLeuGluSer 455
|||||:|||||:|||||:|||||:|||||:|||||:|||||
43 .....CCTCCTGCCCGCA 32
|||||:|||||:|||||:|||||:|||||:|||||:|||||
456 ProSerProSerSerLeuSerProProGlyArg 466
```

seq\_name: /cgn2\_6/ptodata/2/paa/US09\_NEW\_COMB.pep:US-09-935-625-17076

```
seq_documentation_block:
; Sequence 17076, Application US/099335625
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: POLYNUCLEOTIDES, POLYPEPTIDES, CELLS, AND METHODS THEREOF CAPABLE
; FILE REFERENCE: 2750-1481P
; CURRENT APPLICATION NUMBER: US/09/935,625
; CURRENT FILING DATE: 2001-08-24
; NUMBER OF SEQ ID NOS: 33136
; SEQ ID NO 17076
; LENGTH: 847
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: peptide
; LOCATION: 1..847
```

```
; OTHER INFORMATION: Ceres Seq. ID no. 2708207
US-09-935-625-17076
```

```
alignment_scores:
  Quality: 119.00      Length: 111
  Ratio: 2.052        Gaps: 6
  Percent Similarity: 52.252  Percent Identity: 35.135
```

alignment\_block:

US-09-462-480-4/rev x US-09-935-625-17076 ..

Align seg 1/1 to: US-09-935-625-17076 from: 1 to: 847

```
301 AGAAGCCATTGCGAGGACAGCGCTGCTGCTCTCT...CGTCGCC 255
|||||:|||||:|||||:|||||:|||||:|||||:|||||
381 ArgSerProLeuAlaLysArgGlyArgSerAspSerProGlyArgSerPr 397
|||||:|||||:|||||:|||||:|||||:|||||:|||||
254 CTCGAGTATTGGACCGCGGCTGACGAATATTCTCGTGGAGA..... 215
|||||:|||||:|||||:|||||:|||||:|||||:|||||
397 oSerProValAlaArgLeuArgAspProThrGlyAlaArgLeuProSerP 414
|||||:|||||:|||||:|||||:|||||:|||||:|||||
214 .....TCTCGTCGAGTTCTCTGCTTCTGCTTATTGCTGCTCTTTGGA 173
|||||:|||||:|||||:|||||:|||||:|||||:|||||
414 roSerIleGluGlnArgLeuProSerProProValAlaGlnArgLeuPro 430
|||||:|||||:|||||:|||||:|||||:|||||:|||||
172 AGCGCACACCG...CGGCTGGCGCGCTGCCGCCGCCGCCGCCAC 126
|||||:|||||:|||||:|||||:|||||:|||||:|||||
431 SerProProProArgArgAlaGlyLeuProSerProPro..... 443
125 TGGCCCTGCAACGACCTGCGTGCACCTCCACCTGGTCTGGTGGTTT 76
|||||:|||||:|||||:|||||:|||||:|||||:|||||
444 ....ProAlaGlnArgLeuProSerProProProArgArgAlaGlyLeuP 459
|||||:|||||:|||||:|||||:|||||:|||||:|||||
75 CAGGTCCCGCGGAGA.....TCCGCTCGAAATTACCTG..... 44
|||||:|||||:|||||:|||||:|||||:|||||:|||||
459 roSerProMetArgIleGlyGlySerHisAlaAlaAsnHisLeuGluSer 475
|||||:|||||:|||||:|||||:|||||:|||||:|||||
43 .....CCTCCTGCCCGCA 32
|||||:|||||:|||||:|||||:|||||:|||||:|||||
476 ProSerProSerSerLeuSerProProGlyArg 486
```

seq\_name: /cgn2\_6/ptodata/2/paa/US09\_NEW\_COMB.pep:US-09-935-625-17480

```
seq_documentation_block:
; Sequence 17480, Application US/099335625
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: POLYNUCLEOTIDES, POLYPEPTIDES, CELLS, AND METHODS THEREOF CAPA
; FILE REFERENCE: 2750-1481P
; CURRENT APPLICATION NUMBER: US/09/935,625
; CURRENT FILING DATE: 2001-08-24
; NUMBER OF SEQ ID NOS: 33136
; SEQ ID NO 17480
; LENGTH: 859
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: peptide
; LOCATION: 1..859
; OTHER INFORMATION: Ceres Seq. ID no. 3023746
US-09-935-625-17480
```

```
alignment_scores:
  Quality: 119.00      Length: 111
  Ratio: 2.052        Gaps: 6
  Percent Similarity: 52.252  Percent Identity: 35.135
```

alignment\_block:

US-09-462-480-4/rev x US-09-935-625-17480 ..



```

Align seg 1/1 to: US-09-935-625-17480 from: 1 to: 859

301 AGAAGCCATTTCGAGGACAGCGCTGCTGCTGCTCCT...CGTCGGCC 255
|||||:||||:||||:||||:||||:||||:||||:||||:
393 ArgSerProLeuAlaArgGlyArgSerAspSerProGlyArgSerPr 409
|||||:||||:||||:||||:||||:||||:||||:||||:
254 CTCGAGTATTGGACGCGCGCTCAGCAATATTTCGTCGAGA..... 215
|||||:||||:||||:||||:||||:||||:||||:||||:
409 oSerProValAlaArgLeuArgAspProThrGlyAlaArgLeuProSerP 426
|||||:||||:||||:||||:||||:||||:||||:||||:
214 .....TCTCGTCGAGTTCCTGCTCTCTGCTATTGGCTGCTCTTGGGA 173
|||||:||||:||||:||||:||||:||||:||||:||||:
426 roSerIleGluGlnArgLeuProSerProProValAlaGlnArgLeuPro 442
|||||:||||:||||:||||:||||:||||:||||:||||:
172 AGCGCACCAACCG...CGGCCTGGCGGCGCGTCCCGCGCGCGCGCCAC 126
|||||:||||:||||:||||:||||:||||:||||:||||:
443 SerProProArgArgAlaGlyLeuProSerProPro..... 455
|||||:||||:||||:||||:||||:||||:||||:||||:
125 TGGCCCTCGACGAACTCCCTCGACTCCACTGGTCGATCGGTGGTTT 76
|||||:||||:||||:||||:||||:||||:||||:||||:
456 ...ProAlaGlnArgLeuProSerProProProArgArgAlaGlyLeuP 471
|||||:||||:||||:||||:||||:||||:||||:||||:
75 CAGGTCCCGCGGAGA.....TCGCGTCGAAATTACCTG..... 44
||| ||| |||:||||:||||:
471 roSerProMetArgIleGlyGlySerHisAlaAlaAsnHisLeuGluSer 487
|||||:||||:||||:||||:||||:||||:||||:||||:
43 .....CCTCCTGCCCGCA 32
|||||:||||:||||:||||:||||:||||:||||:||||:
488 ProSerProSerSerLeuSerProProGlyArgArq 498
|||||:||||:||||:||||:||||:||||:||||:||||:

```

THIS PAGE BLANK (USPTO)





803 .....ProProAlaGlySer.....AlaLeuGlyAla 812  
166 CCACCGCGCTGGCGCGCTCCCGCGCGCGCGCGCGCTGC 117  
813 ProProValProSerArgProGlyAlaSerProAspProPheGlyProPr 829  
116 AACGAACCTGCCGTCG.....ACTCCACCTGGTCGATCTGGGT 79  
829 oProGlnValProSerArgProAsnArgAlaProProGlyValProSerA 846  
78 TTTCAGCTCGCGGAGATCGCTCGAATAATACCTGCCT 41  
846 rgSerGlyGlnAlaSerProSerArgProGluSerPro 858

seq\_name: pir1:EDBEIF

seq\_documentation\_block:  
immediate-early protein IE180 - suid herpesvirus 1 (strain Indiana-Funkhauser)  
C:Species: suid herpesvirus 1  
C:Date: 30-Jun-1990 #sequence\_revision 30-Jun-1990 #text\_change 16-Feb-1997  
C:Accession: S04713  
R:Cheung, A.K.

Nucleic Acids Res. 17, 4637-4646, 1989  
A:Title: DNA nucleotide sequence analysis of the immediate-early gene of pseudorabies vi  
A:Reference number: S04713; MUID:89315207  
A:Accession: S04713  
A:Molecule type: DNA  
A:Residues: 1-1460 <CHE>  
C:Superfamily: herpesvirus immediate-early protein IE175  
C:Keywords: DNA binding; early protein; transcription regulation

alignment\_scores:  
Quality: 104.00 Length: 154  
Ratio: 1.926 Gaps: 5  
Percent Similarity: 35.065 Percent Identity: 25.325

alignment\_block:  
US-09-462-480-4/rev x EDBEIF ..

Align seg 1/1 to: EDBEIF from: 1 to: 1460

298 AGCCATTTCGAGGACAGCGCCCTGCTGCTCCTCGTGGCGCCCTGGAG 249  
115 SerProAlaAlaGlySerProValGlyLeuSerIleArgAlaProSerTh 131  
248 TATT.....GGAC 241  
131 rValThrSerSerSerglyProGlyProAlaProGlyProGlyA 148  
240 GCGGCTGAGCAATATTCTGTCGAGATCTGTCGAGTTCCTGCTTCGCT 191  
148 rArgProArgGlnHisSerGlnArgGlnArgProGlyProProAlaAla 164  
190 TATTGGCTGCTTCTTGAAGCGCACCGCGGCTGGCGGCGCTGCC 141  
165 ProGly.....AlaArgProProGlnProProArgProProPr 178  
140 GCCG..... 137  
178 qProAlaProProAlaProProAlaProProAlaProArgProA 195  
137 ..... 137  
195 rGlyAspGlyProProArgGlyGlyThrArgSerValSerProGlyArg 211  
136 GCGCGCGCCACTGGCTGCAACGAAC..... 110  
212 ArgArgGlyLeuGlyProArgArgHisGlnHisSerGlnGlnArgTrpPr 228  
109 .....CTGCCGTCGACTCCACCTG 91  
|||||

228 oGlnArgArgHisGlyGlyProLeuProGlnProProProProProg 245  
90 GTGATCTGGGTTTTCAGGTCCGCGAGATCCGCTCGAAATACCTGCCT 41  
245 lyArgSer.....ArgArgProAlaAlaAlaAlaProPro 256  
40 CCTGCCCGAGGG 29  
257 ProAlaGluGly 260

seq\_name: pir2:T35785

seq\_documentation\_block:  
probable beta-glucosidase - Streptomyces coelicolor  
C:Species: Streptomyces coelicolor  
C:Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 21-Jan-2000  
C:Accession: T35785  
R:Seeger, K.J.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, July 1998  
A:Reference number: Z21570  
A:Accession: T35785  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-859 <SEE>  
A:Cross-references: EMBL:AL031013; PIDN:CAA19790.1; GSPDB:GN00070; SCOEDB:SC8A6.18  
A:Experimental source: strain A3(2)  
C:Genetics:  
A:Gene: SCOEDB:SC8A6.18  
C:Superfamily: beta-glucosidase

alignment\_scores:

Quality: 99.50 Length: 92  
Ratio: 2.117 Gaps: 5  
Percent Similarity: 51.087 Percent Identity: 38.043

alignment\_block:  
US-09-462-480-4 x T35785 ..

Align seg 1/1 to: T35785 from: 1 to: 859

39 GGAGGACAGTAATTCGAGCGGATCTCCGGGACCTGAAACCCAGATCG 88  
287 GlyGlyArg.....AspValSerArgSerProValProGlyAr 299  
89 ACCAGGT.....GGAGTCGACGCGAGGTTCTGTGCAG 120  
299 gProAlaAspSerProGlyAlaGluGlyAlaAspGlyGlyAlaG 316  
121 GCCCAGTGGCGCGCGCGGCGGACGCCGCCCGCGGT..... 164  
316 lyProSerSer.....GlyAlaGluGlyLeuProGlyArgGlyProAla 330  
165 ...GGTCGGCTTCCAGAAGCAGCCCAATAGCAGAGCAAGCAACTCAGC 211  
331 HisGlyAlaLysProSerGlyProArgProArgAlaGlyAspGlyAr 347  
212 AGATCTCGACGAATATTCTGTCAGCGCGGTCCTCAATACTCGAGCGCCGAC 261  
347 gAlaLeu.....AlaArgA 352  
262 GAGGACGAGCAGCAGCGGCTGTCCTC 287  
352 rgAlaValAlaAlaGlyAlaValLeu 360

seq\_name: pir2:B75514

seq\_documentation\_block:  
penicillin-binding protein 1 - Deinococcus radiodurans (strain R1)  
C:Species: Deinococcus radiodurans  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 31-Mar-2000  
C:Accession: B75514  
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J

M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999  
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A:Reference number: A75250; MUID:20036896  
A:Accession: B75514  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-873 <WHI>  
A:Cross-references: GB:AE001907; GB:AE005513; NID:g6458162; PIDN:AAF10059.1; PID:g645816  
A:Experimental source: strain R1  
C:Genetics:  
A:Gene: DR0479  
A:Map position: 1

alignment\_scores:  
Quality: 99.50 Length: 135  
Ratio: 1.746 Gaps: 9  
Percent Similarity: 42.22 Percent Identity: 32.593

alignment\_block:  
US-09-462-480-4/rev x B75514

Align seg 1/1 to: B75514 from: 1 to: 873

```
298 AGCCATTTCGAGGAGCAGCCCTGCTGCTCTGCTCGCGCCCTCGAG 249
      :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
724 ThrProAlaValThrAlaPro.....GluProGlnProAlaTh 738
248 T.....ATTGAGCGCGGCTGAGGATATTCGTCGAGATCTGTC 208
      :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
738 ThrThrProAlaGlyArgSerProAlaProLeuArgArgThrArgA 755
207 GAGTT.....CCTGCTTCTGCTTATTGGCTGCTCTT 176
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
755 TgThrLeuProArgGlnThrSerProArgArgAla.....Leu 767
175 GGAAGCGCACCCAGCGGCTGCGCGCGG...TCCCGCGCG..... 137
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
768 ProSerArgThrCysArgProArgProLeuSerProProArgArg 784
136 .....GCAACGAACCTCGCGTCG..... 101
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
801 roSerAlaThrArgCysProThrCysProArgArgLeuSerArgCys 817
100 .....ACTCCACCTGGTCG 87
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
818 ProAlaThrProArgProLeuSerArgArgProValThrProGlyAr 834
86 ATCTGGTTTTCAGTCGCGGAGATCGCTCGAAATACCTGCTCTG 37
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
834 gPro...PheGlnAlaArgArgGilleProArgLeuArgilePro 850
36 CCGCA 32
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
850 hrArg 851
```

seq\_name: pir2:T18535

seq\_documentation\_block:  
high molecular mass nuclear antigen - chicken (fragment)  
C:Species: Gallus gallus (chicken)  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T18535  
R:Shimada, K.; Harata, M.; Mizuno, S.  
J. Cell Sci. 110, 3031-3041, 1997  
A:Title: A nuclear matrix-associated high molecular mass nuclear antigen, HMNA, of chick  
A:Reference number: Z18955; MUID:9803440

A:Accession: T18535  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: mRNA  
A:Residues: 1-1151 <SHI>  
A:Cross-references: EMBL:D88440; NID:d1177138; PID:d1025045; PIDN:BAA24137.1

alignment\_scores:

Quality: 99.00 Length: 115  
Ratio: 2.062 Gaps: 4  
Percent Similarity: 41.739 Percent Identity: 26.957

alignment\_block:  
US-09-462-480-4/rev x T18535

Align seg 1/1 to: T18535 from: 1 to: 1151

```
298 AGCCATTTCGAGGAGCAGCCCTGCTGCTCTGCTCGCGCCCTCGAG 249
      :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
16 AlaProSerProAlaProAlaProAlaProAlaProAlaPro..... 30
248 TATTGGAGCGCGGCTGACGAATATTCGTCGAGATCTGTCGAGTTC 199
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
31 .....ProArgProLysTrpValProI 38
198 CTTCGCTTAT.....TGGCTGCTT 179
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
38 leAlaGluLeuHisProAlaAlaProGlnProProLysTrpValPro 54
178 CTTGGAAGCGCACCCAGCGGCTGGG...CGCGCTCCCGCGCGCGCG 132
      :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
55 IleGlyGlyAlaProProProGlyThrGluProThrProProSerLy 71
131 C.....GCCACTG 124
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
71 sProThrAspGlyAlaAspAlaAlaProLysAlaSerAlaGluLeuThrS 88
123 GCCCTGCAACGAACCTGCGCTGACTCCACCTGCTGATCTGGGTTTCA 74
      :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
88 erProProAlaSerProSerProProAspGlyProLysAlaProSer 104
73 GGTCCGCGGAGATCCGCTCGAAATACCTGCTCTGCGCGGAGG 29
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
105 GlyAlaGlyGluAlaGluAlaGlyThrProProProSerGlnGly 119
```

seq\_name: pir2:T16953

seq\_documentation\_block:  
hypothetical protein T28D9.2 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 21-Jan-2000  
C:Accession: T16953  
R:Fulton, L.  
submitted to the EMBL Data Library, June 1995  
A:Description: The sequence of C. elegans cosmid T28D9.  
A:Reference number: Z18614  
A:Accession: T16953  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-208 <FUL>  
A:Cross-references: EMBL:U28738; NID:g861262; PID:g861269; PIDN:AAA68314.1; CESP:T28D  
A:Experimental source: strain Bristol N2  
C:Genetics:  
A:Gene: CESP:T28D9.2  
A:Introns: 63/2; 147/2; 173/3  
C:Superfamily: unassigned ribonucleoprotein repeat-containing proteins; ribonucleopro  
F:3-62/Domain: ribonucleoprotein repeat homology <RRM>

alignment\_scores:

Quality: 97.50 Length: 70  
Ratio: 2.267 Gaps: 3  
Percent Similarity: 61.429 Percent Identity: 42.857











OM of: US-09-462-480-4 to: SwissProt\_40:\* out\_format : pfs  
Date: Jul 22, 2002 1:41 AM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

## Command line parameters:

-MODEL=framet\_n2p.model -DEV=xlh  
-O=/cfn2.1/USPTO\_spo01/US09462480/runat\_18072002\_164420\_19600/app\_query.fasta\_1.2850  
-DB=SwissProt\_40 -QFM=fastan -SUFFIX=esp -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000  
-CGAPOP=4.000 -CGAPOP=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-FCGAPOP=6.000 -FCGAPOP=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09462480.ecgn1\_1.29 -NCPU=6 -ICPU=3 -LONLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

## Search information block:

Query: US-09-462-480-4

Query length: 302

Database: SwissProt\_40:\*

Database sequences: 105224

Database length: 38719550

Search time (sec): 49.760000

## score\_list:

Sequence	Strd	Orig	zScore	Escore	Len
SwissProt_40:CF10_MYCTU	+	481.00	710.03	7.6e-32	99
SwissProt_40:CF10_MYCLE	+	182.00	274.67	1.3e-07	99
SwissProt_40:DYN1_HUMAN	+	105.50	146.80	0.2040	864
SwissProt_40:YX2_CAEEL	+	97.50	145.99	0.9410	208
SwissProt_40:DYN1_RAT	-	97.50	135.27	0.9092	851
SwissProt_40:DYN1_MOUSE	-	97.50	135.18	0.9089	861
SwissProt_40:Y135_STRFR	-	93.50	136.25	1.96	348
SwissProt_40:Y135_PVKA	-	93.50	125.41	1.89	1446
SwissProt_40:CUT1_CANFA	-	93.00	127.68	2.10	975
SwissProt_40:CBPA_DICDI	-	92.00	131.83	2.58	467
SwissProt_40:WAIP_HUMAN	-	91.50	130.53	2.82	503
SwissProt_40:TAU_MOUSE	-	91.50	127.68	2.80	732
SwissProt_40:CC19_CAEEL	-	90.50	133.45	3.45	283
SwissProt_40:WAS3_HUMAN	-	90.00	128.36	3.74	502
SwissProt_40:NNE3_RAT	-	90.00	121.50	3.66	1237
SwissProt_40:GSRI_HUMAN	-	90.00	119.99	3.64	1509
SwissProt_40:ROK_HUMAN	-	89.50	128.25	4.11	463
SwissProt_40:ROK_RABIT	-	89.50	128.25	4.11	463
SwissProt_40:ROK_MOUSE	-	89.50	128.24	4.11	464
SwissProt_40:WASP_HUMAN	-	89.50	127.64	4.10	502
SwissProt_40:ES6D_MYCTU	+	89.00	138.96	4.68	103
SwissProt_40:IRS2_HUMAN	-	89.00	119.53	4.40	1324
SwissProt_40:WS14_MOUSE	-	88.50	122.05	4.88	864
SwissProt_40:IF2_MYCLE	+	88.00	120.81	5.35	924
SwissProt_40:ANT1_ONCVO	+	87.00	113.40	6.33	2022
SwissProt_40:S3A2_MOUSE	-	86.00	123.14	7.90	464
SwissProt_40:S3A2_MOUSE	-	86.00	122.96	7.90	475
SwissProt_40:RP3A_BOVIN	-	85.50	119.24	8.59	704
SwissProt_40:IF2_MYCTU	+	85.50	117.37	8.54	900
SwissProt_40:VGLB_HSVBP	-	85.50	117.14	8.53	928
SwissProt_40:VGLB_HSVBC	-	85.50	117.11	8.53	932
SwissProt_40:SFR2_CHICK	+	85.00	127.33	9.70	221
SwissProt_40:ICP3_HSVIF	+	85.00	126.00	9.66	263
SwissProt_40:SYN1_HUMAN	-	84.50	117.77	10.35	705
SwissProt_40:TRX2_HUMAN	-	84.50	107.52	10.02	2715
SwissProt_40:YML2_THFE	-	84.00	131.32	11.90	108
SwissProt_40:ICP3_HSVIF	-	84.00	124.87	11.65	252
SwissProt_40:NIF1_MOUSE	-	84.00	122.42	11.56	348
SwissProt_40:CCPA_ACEXY	-	84.00	122.31	11.56	353
SwissProt_40:T2D3_HUMAN	-	84.00	113.78	11.25	1083

SwissProt\_40:TAU\_RAT - 83.50 115.84 12.46 751 ! P19332 rattus norvegicus (rat)  
SwissProt\_40:MYSD\_DICDI - 83.00 112.12 13.55 1113 ! P34109 dictyostellium discoidi  
SwissProt\_40:ICPA\_HVMG - 83.00 110.29 13.47 1415 ! Q02362 marek's disease 1  
SwissProt\_40:VNUA\_PVKA - 83.00 108.75 13.40 1733 ! P33485 pseudorabies virus  
SwissProt\_40:SCE3\_SCHPO - 82.50 119.40 15.26 388 ! O14369 schizosaccharomyces  
seq\_name: SwissProt\_40:CF10\_MYCTU  
seq\_documentation\_block:  
ID CF10\_MYCTU STANDARD; PRT; 99 AA.  
AC Q69739;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE 10 kDa culture filtrate antigen cfp10.  
GN CFP10 OR LHP OR RV3874 OR MT3988 OR MTV027.09.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-15.  
RC STRAIN=H37RV;  
RX MEDLINE=99061212; PubMed=9846755;  
RA Berthet F.X., Rasmussen P.B., Rosenkrands I., Andersen P.,  
RA Gicquel B.;  
RT "A Mycobacterium tuberculosis operon encoding ESAT-6 and a novel  
RT low-molecular-mass culture filtrate protein (CFP-10).";  
RL Microbiology 144:3195-3203(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H37RV;  
RX MEDLINE=98295987; PubMed=9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies K., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
RT complete genome sequence.";  
RL Nature 393:537-544(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CDC 1551 / Oshkosh;  
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
RA Bishai W.;  
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
RT laboratory strains.";  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: BELONGS TO THE ESAT6 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL: AF004671; AAC83445.1; -  
DR EMBL: AL022120; CAA17966.1; -  
DR EMBL: AE007190; AAK48356.1; -  
TIGR: MT3988; -  
DR TIGR: MT3988; -  
DR TuberculList: RV3874; -  
KW Complete proteome.  
INIT\_MET 0  
SQ SEQUENCE 99 AA; 10663 MW; EBCA6GA996C5489D CRC64;

alignment\_scores:  
Quality: 481.00 Length: 99  
Ratio: 4.859 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 98.990

alignment\_block:  
US-09-462-480-4 x CF10\_MYCTU

Align seg 1/1 to: CF10\_MYCTU from: 1 to: 99

4 GCAGAGATGAAGACCGGCTACCTCGGCGAGGAGGAGGTAATTT 53  
1 AlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAsnPh 17

54 CGAGCGGATCTCCGGGACCTGAACCCAGATCGACGAGTGGAGTGA 103  
17 eGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGluSer 34

104 CGGCGAGTTCGTTCAGGCGCAGTGGCGCGGCGGCGGAGCGCGCC 153  
34 hrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAlaAla 50

154 CAGGCGCGGTGGTGGCTTCCAGAGCAGCCAAATAGCAGAGCAGGA 203  
51 GlnAlaAlaValValArgPheGlnGluAlaAlaAsnLysGlnLysGln 67

204 ACTCGACGAGATCTCGACGAATATTCGTTCAGGCGCGGCTCCAA 253  
67 uLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTrpSer 84

254 GGGCGCAGCAGGAGCAGCAGGAGCGCTCTCTCGCAATGGGCTTC 300  
84 rgAlaAspGluGlnGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 99

seq\_name: SwissProt\_40:CF10\_MYCLE

seq\_documentation\_block:  
ID CF10\_MYCLE STANDARD; PRT; 99 AA.

AC 033084;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE 16 kDa culture filtrate antigen cfpl0 homolog.  
GN M0050 OR MLCB628.13C.  
OS Mycobacterium leprae.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1769;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Cole S.T., Eiglmeier K., Garnier T., De Rossi E., Fsihi H., Cole S.T.;  
RA Eiglmeier K., Garnier T., De Rossi E., Fsihi H., Cole S.T.;  
RA Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
RL [2]  
RN SEQUENCE FROM N.A.  
RC STRAIN=TN;  
RX MEDLINE=21128732; PubMed=11234002;  
RA Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,  
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,  
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,  
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,  
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,  
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,  
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,  
RA Barrell B.G.;  
RT "Massive gene decay in the leprosy bacillus";  
RL Nature 409:1007-1011(2001).  
CC -!- SIMILARITY: BELONGS TO THE ESAT6 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; Y14967; CAA75210.1;  
DR EMBL; AL583917; CAC29558.1;  
DR Leproma; M0050;  
KW Complete proteome.  
FT INIT\_MET 0 0 BY SIMILARITY.  
SQ SEQUENCE 99 AA; 10833 MW; 859B484F7EFE5A8A CRC64;

alignment\_scores:  
Quality: 182.00 Length: 99  
Ratio: 2.333 Gaps: 0  
Percent Similarity: 78.788 Percent Identity: 39.394

alignment\_block:  
US-09-462-480-4 x CF10\_MYCLE

Align seg 1/1 to: CF10\_MYCLE from: 1 to: 99

4 GCAGAGATGAAGACCGGCTACCTCGGCGAGGAGGAGGTAATTT 53  
1 AlaGluMetIleThrGluAlaAlaIleLeuThrGlnGlnAlaAlaGlnPh 17

54 CAGGCGGATCTCCGGGACCTGAACCCAGATCGACGAGTGGAGTGA 103  
17 eAspGlnIleAlaSerGlyLeuSerGlnGluArgAsnPheValAspSerI 34

104 CGGCGAGTTCGTTCAGGCGCAGTGGCGCGGCGGCGGAGCGCGCC 153  
34 leglyGlnSerPheGlnAsnThrTrpGluGlyGlnAlaAlaSerAlaAla 50

154 CAGGCGCGGTGGTGGCTTCCAGAGCAGCCAAATAGCAGAGCAGGA 203  
51 LeuGlyAlaLeuGlyArgPheAspGluAlaMetGlnAspGlnIleArgG 67

204 ACTCGACGAGATCTCGACGAATATTCGTTCAGGCGCGGCTCCAA 253  
67 nLeuGluSerIleValAspLysLeuAsnArgSerGlyGlyAsnTrpThrL 84

254 GGGCGCAGCAGGAGCAGCAGGAGCGCTCTCTCGCAATGGGCTTC 300  
84 yThrAspAspGluAlaAsnGlnLeuLeuSerSerLysMetAsnPhe 99

seq\_name: SwissProt\_40:DYN1\_HUMAN

seq\_documentation\_block:  
ID DYN1\_HUMAN STANDARD; PRT; 864 AA.

AC 005193;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Dynamin-1 (BC 3.6.1.50).  
GN DNM1 OR DNM  
OS Homo sapiens (Human)  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93328758; PubMed=8101525;  
RA van der Bliek A.M., Redelmeier T.E., Tisdale E.J., Meyerowitz E.M.,  
RA Schmid S.L.;  
RT "Mutations in human dynamin block an intermediate stage in coated  
RT vesicle formation";  
RL J. Cell Biol. 122:553-563(1993).  
RN [2]  
RP STRUCTURE BY NMR OF 511-630.  
RX MEDLINE=95153276; PubMed=7850421;

RA Downing A.K., Driscoll P.C., Gout I., Salim K., Zvelebil M.J.,  
RA Waterfield M.D.;  
RT "Three-dimensional solution structure of the pleckstrin homology  
RT domain from dynamin.";  
RL Curr. Biol. 4:884-891(1994).  
RN [3]  
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 509-633.  
RX MEDLINE=95042718; PubMed=7954789;  
RA Ferguson K.M., Lemmon M.A., Schlessinger J., Sigler P.B.;  
RT "Crystal structure at 2.2-A resolution of the pleckstrin homology  
RT domain from human dynamin.";  
RL Cell 79:199-209(1994).  
RN [4]  
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 509-630.  
RX MEDLINE=95360729; PubMed=7634088;  
RA Tilm D., Salim K., Gout I., Guruprasad L., Waterfield M., Blundell T.;  
RT "Crystal structure of the pleckstrin homology domain from dynamin.";  
RL Nat. Struct. Biol. 1:782-788(1994).  
CC -|- FUNCTION: MICROTUBULE-ASSOCIATED FORCE-PRODUCING PROTEIN INVOLVED  
CC IN PRODUCING MICROTUBULE BUNDLES AND ABLE TO BIND AND HYDROLYZE  
CC GTP. MOST PROBABLY INVOLVED IN VESICULAR TRAFFICKING PROCESSES, IN  
CC PARTICULAR ENDOCYTOSIS.  
CC -|- CATALYTIC ACTIVITY: GTP + H(2)O = GDP + phosphate.  
CC -|- SUBCELLULAR LOCATION: MICROTUBULE-ASSOCIATED.  
CC -|- SIMILARITY: BELONGS TO THE DYNAMIN FAMILY.  
CC -|- SIMILARITY: CONTAINS 1 PH DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; L07807; AAA02803.1; -;  
DR EMBL; L07808; AAA02804.1; ALT\_SEQ.  
DR EMBL; L07809; AAA02805.1; ALT\_SEQ.  
DR EMBL; L07810; AAA02806.1; ALT\_SEQ.  
DR PDB; 1DXN; 27-FEB-95.  
DR PDB; 2DXN; 12-NOV-97.  
DR MIM; 602377; -;  
DR InterPro; IPR001401; Dynamin.  
DR InterPro; IPR000375; Dynamin\_central.  
DR InterPro; IPR003130; GED.  
DR InterPro; IPR001849; PH.  
DR Pfam; PF00350; dynamin; 1.  
DR Pfam; PF01031; dynamin\_2; 1.  
DR Pfam; PF02212; GED; 1.  
DR Pfam; PF00169; PH; 1.  
DR PRINTS; PR00195; DYNAMIN.  
DR SMART; SM00053; DYNC; 1.  
DR SMART; SM00302; GED; 1.  
DR SMART; SM00233; PH; 1.  
DR PROSITE; PS00410; DYNAMIN; 1.  
DR PROSITE; PS50003; PH\_DOMAIN; 1.  
KW Hydrolyase; Motor protein; GTP-binding; Microtubules; Multigene family;  
KW Endocytosis; 3D-structure.  
FT NP\_BIND 38 45 GTP (BY SIMILARITY).  
FT NP\_BIND 136 140 GTP (BY SIMILARITY).  
FT NP\_BIND 205 208 GTP (BY SIMILARITY).  
FT DOMAIN 519 625 PH.  
SQ SEQUENCE 864 AA; 97407 MW; 7B2D86B572FE0EF CRC64;

alignment\_scores:  
Quality: 105.50 Length: 96  
Ratio: 2.198 Gaps: 4  
Percent Similarity: 50.000 Percent Identity: 35.417

alignment\_block:  
US-09-462-480-4/rev x DYN1\_HUMAN

Align seg 1/1 to: DYN1\_HUMAN from: 1 to: 864  
301 AGAAGCCCATTTGCG.....AGGACAGCGCTGCTGCTGC 267  
|||||  
773 ArgSerProThrSerSerProThrProGlnArgArgAlaProAlaValPr 789  
|||||  
266 TCCCTGCTGCGCCCTCGAGTAGTATTGGACGCCGCGCTGACGAATATTGCTGA 217  
||| |||||  
789 oProAlaArgProGlySerArgGlyProAlaProGlyPro..... 802  
216 GATCTGCTGAGTTCCTGCTCTCTGCTTATTGCTGCTCTCTTGGAGCGCA 167  
|||||  
803 .....ProProAlaGlySer.....AlaLeuGlyGlyAla 812  
166 CCACCGCGGCTGCGGCGGCTGCTGCCGCGCGCGGCGGCGGCGGCGGCTGC 117  
|||||  
813 ProProValProSerArgProGlyAlaSerProAspProPheGlyProPr 829  
116 AACGAACCTGCGCTG.....ACTCCACCTGGTGCATCGTGGT 79  
: : : : :  
829 oProGlnValProSerArgProAsnArgAlaProProGlyValProSerA 846  
78 TTTTCAGTTCGCGGAGATCCGCTCGAAATTACCTGCCT 41  
|||||  
846 rgSerGlyGlnAlaSerProSerArgProGluSerPro 858  
seq\_name: SwissProt\_40:YSX2\_CAEEL  
seq\_documentation\_block:  
ID YSX2\_CAEEL STANDARD; PRT; 208 AA.  
AC Q10021;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical 24.0 kDa protein T28D9.2 in chromosome II.  
OS T28D9.2.  
GN Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RA Waterston R.;  
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.  
CC -|- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U28738; AAA68314.1; -;  
DR WormPep; T28D9.2; CEO2067.  
DR InterPro; IPR000504; RRM.  
DR Pfam; PF00076; rrm; 1.  
DR SMART; SM00360; RRM; 1.  
DR PROSITE; PS50102; RRM; 1.  
DR PROSITE; PS00030; RRM\_RNP\_1; FALSE\_NEG.  
KW Hypothetical protein; RNA-binding.  
FT DOMAIN 2 74 RNA-BINDING (RRM).  
FT DOMAIN 84 205 ARG/SER-RICH.  
SQ SEQUENCE 208 AA; 23992 MW; 0FE8AAD05FB777EF CRC64;

alignment\_scores:  
Quality: 97.50 Length: 70  
Ratio: 2.267 Gaps: 3  
Percent Similarity: 61.429 Percent Identity: 42.857

US-09-462-480-4 X YSX2 CAPEL

\_\_\_\_\_

\_\_\_\_\_







```

CC -|- SIMILARITY: BELONGS TO THE CUT FAMILY OF HOMEBOX PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, Usage by and for Commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X69017; CAA48782.1; -.
CC DR PIR: S33121; S33121.
CC DR HSP: P10037; 1A07.
CC DR TRANSFAC: T01485; -.
CC DR TRANSFAC: T01995; -.
CC DR InterPro: IPR003350; CUT.
CC DR InterPro: IPR001356; Homeobox.
CC DR Pfam: PF02376; CUT; 3
CC DR Pfam: PF00046; homeobox; 1.
CC DR SMART: SM00389; HOX; 1.
CC DR PROSITE: PS00027; HOMEBOX_1; 1.
CC DR PROSITE: PS00071; HOMEBOX_2; 1.
CC KW Transcription regulation; Homeobox; DNA-binding;
CC KW Developmental protein; Nuclear protein; Repeat; Repressor;
CC KW Coiled coil; Alternative splicing.
CC -----
CC NON_TER 1 1
CC FT DNA_BIND <1 67 CUT 1. COIL (POTENTIAL).
CC FT DOMAIN 113 169 COILED
CC FT DNA_BIND 374 461 CUT 2.
CC FT DNA_BIND 557 644 CUT 3.
CC FT DNA_BIND 684 743 HOMEBOX.
CC SQ SEQUENCE 975 AA; 105428 MW; 724243632C5BCFD9 CRC64;

```

[illegible]

```

|||||
904 yAlaHisGargSerAlaLeuPro 912

seq_name: SwissProt_40:CBPA_DICDI
seq_documentation_block:
ID CBPA_DICDI STANDARD; PRT; 467 AA.
AC F35085;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Calcium-binding protein.
GN CBPA.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetoza; Dictyosteliida; Dictyostellium.
OX NCBI_TaxID=44689;
RN [1]
RS SEQUENCE FROM N.A.
RP STRAIN=AX2;
RC Wenington R., Greenwood M., Tsang A.;
RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
CC -! SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement/ or send an email to licenselib-sib.ch).
-----
CC EMBL; U03413; AAA03471.1; -.
DR DictyDb; DD01055; cbpa.
DR InterPro; IPR002048; EF-hand.
DR Pfam; PF00036; efhand; 2.
DR SMART; SM00054; Eph; 2.
DR PROSITE; PS00018; EF_HAND; 1.
KW Calcium-binding.
FT CA_BIND 412 423 EF_HAND (POTENTIAL).
SQ SEQUENCE 467 AA; 48976 MW; 2D2055D1E344FC18 CRC64;
```

[illegible]

```
:: ||||| ::|||
135 lyGlnTyrProGlnGly 142
seq_name: SwissProt_40:WAIP_HUMAN

seq_documentation_block:
ID WAIP_HUMAN STANDARD; PRT; 503 AA.
AC Q43516; Q9UNP1; Q15220;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Wiskott-Aldrich syndrome protein interacting protein (WASP interacting
DE protein) (PRPL-2 protein).
GN WASPIP OR WIP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=98070810; PubMed=9405671;
RA Ramesh N., Anton I.M., Hartwig J.H., Geha R.S.;
RT "WIP, a protein associated with Wiskott-Aldrich syndrome protein,
RT induces actin polymerization and redistribution in lymphoid cells.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:14671-14676(1997).
RN [2]
SEQUENCE FROM N.A.
RX TISSUE=Tonsil;
RA Kreideweiss S., Delany-Heiken P., Nordheim A., Ruhlmann A.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
SEQUENCE OF 192-503 FROM N.A., AND VARIANT ALA-495.
RX MEDLINE=99218549; PubMed=10202051;
RA Stewart D.M., Tian L., Nelson D.L.;
RT "Mutations that cause the Wiskott-Aldrich syndrome impair the
RT interaction of Wiskott-Aldrich syndrome protein (WASP) with WASP
RT interacting protein.";
RL J. Immunol. 162:5019-5024(1999).
CC -|- FUNCTION: MAY HAVE DIRECT ACTIVITY ON THE ACTIN CYTOSKELETON.
CC -|- INDUCES ACTIN POLYMERIZATION AND REDISTRIBUTION.
CC -|- SUBUNIT: BINDS TO WASP, PROFILIN AND ACTIN.
CC -|- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN PERIPHERAL BLOOD
CC MONONUCLEAR CELLS, SPLEEN, PLACENTA, SMALL INTESTIN, COLON,
CC THYMUS. LOWER EXPRESSION IN OVARY, HEART, BRAIN, LUNG, LIVER,
CC SKELETAL MUSCLE, KIDNEY, PANCREAS, PROSTATE, TESTIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF031588; AAC03767.1; -.
DR EMBL; AF106062; AAD45972.1; -.
DR EMBL; X86019; CAA60014.1; -.
DR MIM; 602357; -.
DR InterPro; IPR003124; WH2.
DR Pfam; PF02205; WH2; 1.
DR SMART; SM00246; WH2; 1.
KW Actin-binding; Repeat.
FT DOMAIN 2 13
FT DOMAIN 64 96
FT DOMAIN 241 244
FT DOMAIN 264 433
FT REPEAT 352 361
FT REPEAT 374 383
FT REPEAT 410 419
FT SITE 45 48
FT VARIANT 495 495
FT G -> A.
FT /FTID=VAR_010295.
FT PHRPLR -> SSQAPPP (IN REF. 3).
FT CONFLICT 303 309

FT CONFLICT 360 360 P -> PV (IN REF. 2).
FT CONFLICT 487 503 GSNRRRGPPPIPR -> EYFCQGF (IN REF. 2).
SQ SEQUENCE 503 AA; 51489 MW; 43EB8674DD3BF1A CRC64;

alignment_scores:
Quality: 91.50 Length: 107
Ratio: 1.906 Gaps: 5
Percent Similarity: 44.860 Percent Identity: 33.645

alignment_block:
US-09-462-480-4/rev x WAIP_HUMAN ..
Align seg 1/1 to: WAIP_HUMAN from: 1 to: 503
277 CCTCGTCTGCTCTCTCGTGGCCCTCGAGTATTGGA..... 242
||||| ||||| ||||| ||||| ||||| |||||
316 ProGlyProProLeuProLeuProSerSerSerGlyAsnAspGluThrPr 332
241 .CGCCGGCCTGACGAATATTCGTCGAGATCTCGAGTCTCTGCTCTG 193
||| ||| ||| ||| ||| ||| ||| ||| |||
332 oArgLeuProGlnArgAsnLeuSerLeuSerSerThrProLeuP 349
192 CTTATTGGCTCTTCTTGGAGCCGACCACCGCGCGCTGGG....CGGCCG 146
||| ||| ||| ||| ||| ||| ||| ||| |||
349 roSerProGlyArgSerGlyProLeuProProProSerGluArgPro 365
145 TCCCGCGCGCGCGC.....GCCACTGGCCCT..... 119
||||| ||| ||| ||| ||| ||| ||| ||| |||
366 ProProValArgAspProProGlyArgSerGlyProLeuProProPr 382
118 .....GCAACGAACCTGCCGTCGACTC 97
382 oProProValSerArgAsnGlySerThrSerArgAlaLeuProAlaThrP 399
96 CA.....CCTGGTCGATCGGGTTTCAGTCCCGGAGATCCGCTCGA 53
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
399 roGlnLeuProSerArgSerGlyValAspSerProArgSerGlyProArg 415
52 AATTACCTGCCTCTGCCCGA 32
||||| ||| |||
416 ProProLeuProProAspArg 422

seq_name: SwissProt_40:TAU_MOUSE

seq_documentation_block:
ID TAU_MOUSE STANDARD; PRT; 732 AA.
AC P10637; P10638; Q60684; Q60685; Q62286;
DT 01-JUL-1989 (Rel. 11, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Microtubule-associated protein tau (Neurofibrillary tangle protein)
DE (Paired helical filament-tau) (PHF-tau).
GN MAPT OR MTAPT OR TAU.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A. (ISOFORM PNS-TAU).
RX TISSUE=Neuroblastoma;
RX MEDLINE=92262443; PubMed=1374898;
RA Couchie D., Mavilia C., Georgieff I.S., Liem R.K.H., Shelanski M.L.,
RA Nunez J.;
RT "Primary structure of high molecular weight tau present in the
RT peripheral nervous system.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:4378-4381(1992).
RN [2]
SEQUENCE FROM N.A. (ISOFORMS TAU-A; TAU-D AND TAU-E).
RX STRAIN=Him OF1; TISSUE=Liver, Kidney, and Brain;
RX MEDLINE=95012085; PubMed=7927211;
RA Kenner L., el-Shabrawi Y., Hutter H., Forstner M., Zatloukal K.,
RA Hoefler G., Preisegger K.-H., Kurzbaue R., Denk H.;
```



```
465 sThrThrProSerProLysThrProGlySerGlyGluProProLys 482
88 .....CGATCGGGTTTCAGTCCCGGAGATCGCTCGAAATAC 47
482 erGlyGluArgSerGlyTyrSerProGlySer..... 493
46 CTGCCTCTCCGCGAGGTAGCGGATCGGTCCTTCATCTCTGCCA 2
494 ...ProGlyThrProGlySerArgSerArgThrProSerLeuPro 507
```

seq\_name: SwissProt\_40:CC19\_CAEEL

seq\_documentation\_block:

```
ID CC19_CAEEL STANDARD; PRT; 283 AA.
AC P18835;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-JUN-1994 (Rel. 19, Last annotation update)
DE Cuticle collagen 19.
GN COL-19.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=89326131; PubMed=2753356;
RA Cox G.N., Fields C., Kramer J.M., Rosenzweig B., Hirsh D.;
RT "Sequence comparisons of developmentally regulated collagen genes of
  Caenorhabditis elegans.";
RL Gene 76:331-344 (1989).
CC -!- FUNCTION: NEMATODE CUTICLES ARE COMPOSED LARGELY OF COLLAGEN-LIKE
CC PROTEINS. THE CUTICLE FUNCTIONS BOTH AS AN EXOSKELETON AND AS A
CC BARRIER TO PROTECT THE WORM FROM ITS ENVIRONMENT.
CC -!- SUBUNIT: COLLAGEN POLYPEPTIDE CHAINS ARE COMPLEXED WITHIN THE
CC CUTICLE BY DISULFIDE BONDS AND OTHER TYPES OF COVALENT CROSS-
CC LINKS.
CC -!- SIMILARITY: TO OTHER COLLAGENS. STRONG, TO OTHER CUTICLE
CC COLLAGENS. COL-7, COL-8 AND COL-19 BELONGS TO THE SAME GROUP OF
CC COLLAGEN.
```

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

```
DR EMBL; M25481; AAA27987.1; -
DR PIR; JS0170; JS0170.
DR InterPro; IPR002486; Col_cuticle_N.
DR Pfam; PF01391; Collagen; 2.
DR Pfam; PF01484; Col_cuticle_N; 1.
KW Cuticle; Connective tissue; Repeat; Multigene family; Collagen.
FT DOMAIN 95 124 TRIPLE-HELICAL REGION.
FT DOMAIN 141 263 TRIPLE-HELICAL REGION.
SQ SEQUENCE 283 AA; 27565 MW; FD53BE2785D1767C CRC64;
```

alignment\_scores:  
Quality: 90.50 Length: 82  
Ratio: 1.847 Gaps: 6  
Percent Similarity: 59.756 Percent Identity: 36.585

alignment\_block:

US-09-462-480-4/rev x CC19\_CAEEL ..

Align seg 1/1 to: CC19\_CAEEL from: 1 to: 283

```
277 CCTGCTGCTCTCTCGCGCCTCGAGTATTGGACGCCGCGCTGACGA 228
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
178 ProGlnGlyAspAlaGlyArgProGlyAlaAlaGly...ArgProGlyPr 193
227 ATATTGCTGAGATCTCGTCGAGTTCCTGCTCTCTGCTTATTGCTGCTTC 178
: :||| :||| :||| :||| :||| :||| :||| :||| :|||
193 oAlaGlyProArgGlyGlu.....ProGlyThrGluTyr..... 204
177 TTGGAAGCGCACACCGCGGCTGGGCGCGCTGCCCG.....CCGCGC 134
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
205 .....LysProGlyGlnProGlyArgProGlyProGlyProGlyProArg 218
133 CGCGCCACTGCTGCACGAACTGCGCTGCGACTCCACTGCTGCTGATC 84
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
219 GlyGluThrGlyProAlaGlyAsn...ProGlyAla...ProGlyAsnAs 233
83 TGGGTTTTTCAGTCTCGCGGAGATCGCTCGAAATACCTGCTGCTCT 38
:||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
233 pGlyGluAlaGlyLysAsnGlyAsnAlaGlyArgProGlyProPro 248
seq_name: SwissProt_40:WAS3_HUMAN
seq_documentation_block:
ID WAS3_HUMAN STANDARD; PRT; 502 AA.
AC Q9UPY6; O94974;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Wiskott-Aldrich syndrome protein family member 3 (WASP-family
  protein member 3) (Verprolin homology domain-containing protein 3).
GN WASF3 OR WAVE3 OR SCAR3 OR KIAA0900.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99310608; PubMed=10381382;
RA Suetsugu S., Miki H., Takenawa T.;
RT "Identification of two human WAVE/SCAR homologues as general actin
  regulatory molecules which associate with the Arp2/3 complex.";
RL Biochem. Biophys. Res. Commun. 260:296-302(1999).
RN [2]
RP SEQUENCE OF 48-502 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99156230; PubMed=10048485;
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirose M.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XII.
  The complete sequences of 100 new cDNA clones from brain which code
  for large proteins in vitro.";
RL DNA Res. 5:355-364(1998).
RN [3]
RP SEQUENCE OF 48-502 FROM N.A.
RA Machesky L.M., Insall R.H.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: DOWNSTREAM EFFECTOR MOLECULES INVOLVED IN THE
CC TRANSMISSION OF SIGNALS FROM TYROSINE KINASE RECEPTORS AND SMALL
CC GTPASES TO THE ACTIN CYTOSKELETON.
CC -!- SUBUNIT: BINDS ACTIN AND ARP2/3 COMPLEX.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN OVARY AND BRAIN.
CC -!- DOMAIN: BINDS ARP2/3 COMPLEX THROUGH THE C-TERMINAL REGION AND
CC ACTIN THROUGH VERPROLIN HOMOLGY (VPH) DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 WH2 DOMAIN
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/)
```

alignment_scores:		
Quality:	90.00	Length: 69
Ratio:	2.812	Gaps: 2
Percent Similarity:	46.377	Percent Identity: 34.783

Align seg 1/1 to: WAS3\_HUMAN from: 1 to: 502

seq name: SwissProt 40:NME3 RAT

Align seq 1/1 to: NME3 RAT from: 1 to: 1237

301 AGAAGCCCATTTTCGAGGACAGCGCTGTGCTCTCTCGTCGGCCCTC 252  
 ||||:||||:||||| |||| ||| |||||:  
 960 ArgThrProLeuAlaArgAlaPro...GlnProProAlaArgProAl 975  
 251 GAGTATTGACGCCGCCCTGACGAATATTGTCGAGATCTGCTCGAGTTC 202  
 :||: ||| :||: ||| ||| ||| |||  
 975 aThrCysGlyProProLeuProAspValSerArgProSerCysArgHisA 992  
 201 CTGCTCTTGGTTATTGGTCTCTTGTGGAGCGCACACCGCGCGCTGGG 152  
 :||:|||| ||| |||| |||||  
 992 IaSerAspAlaArgTrp.....ProValArgValGly 1002

```
151 CGGCCGTCCTCCG.....CGGCCGCGCCGACTGGCCCTGC 117
   ::: :||||| ||||||
1003 HisGlnGlyProHisValSerAlaSerGluArgAlaLeu...ProG1 1018
   :|||||
116 AACGAACCTGCGCTCGACTCCACCTGGTCGATCTGGGTTTTCAGGTCGCC 67
   :||||| :||||| :|||||
1018 uArgSerLeuLeuProAlaHisCysHisTyrSerSerPheProArgAlaG 1035
   :|||||
66 GGAGATCCGCTCGAAATTACCTGCCT..... 41
   :||||| :|||||
1035 LuArgSerGlyArgProTyrLeuProLeuPheProGluProProGluPro 1051
   :|||||
40 .....CCT..... 38
   |||
1052 AspAspLeuProLeuLeuGlyProGluGlnLeuAlaArgArgGluAlaMe 1068
   :|||||
37 .....GCCCGAGGTAGCGGCATCGGTCTTCATCTCTGC 4
   :||||| :||||| :|||||
1068 tLeuArgAlaAlaTrpAlaArgGlyProArgProArgHisAlaSerLeup 1085
   :|||||
3 CA 2
  ||
1085 ro 1085
```



```

75 CAGGTGCGCGGAGA.....TCGTCGAAATTACCTG..... 44
   ::      |||      |||      |||      |||
506 roSerProMetArgIleGlySerHisAlaAlaAsnHisLeuGluSer 522
   43 .....CCTCTGCGCGGA 32
   |||||:::||||
523 ProSerProSerLeuSerProProGlyArg 533
seq_name: sp_human:O94932

seq_documentation_block:
ID   O94932      PRELIMINARY;          PRT;   645 AA.
AC   O94932;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   KIAA0847 PROTEIN (FRAGMENT).
GN   KIAA0847.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
RX   []
RN   SEQUENCE FROM N.A.
RP   TISSUE=BRAIN;
RX   MEDLINE=99156230; PubMed=10048485;
RA   Nagase T., Ishikawa K., Suyama M., Kikuno R., Hirose M.,
RA   Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes.
RT   XII. The complete sequences of 100 new cDNA clones from brain which
RT   code for large proteins in vitro.";
RL   DNA Res. 5:355-364(1998).
DR   EMBL; AB020654; BAA74870.1; -.
SQ   SEQUENCE 1
FT   NON_TER 1
SQ   SEQUENCE 645 AA; 70370 MW; 282EE9ABAB37ADFE CRC64;

```

```

alignment_scores:
  Quality: 109.00      Length: 106
  Ratio: 1.817        Gaps: 3
  Percent Similarity: 56.604  Percent Identity: 30.189

alignment_block:
  US-09-462-480-4/rev x 094932  ..

  Align seg 1/1 to: 094932 from: 1 to: 645

301 AGAAGCCCATTTCCGAGACAGCCCTGCTGCTCTCGTGGCCCTC 252
   :||||||   ::   |||   ::   |||:::   |||
530 LysSerProGlySerProHisAsnProLysThrProLysSerPro.. 545

251 GAGTATTGACGCCGGCCCTGACGAATATTGCTCGAGATCTCGTGATC 202
   ::   |||||   ::   |||   |||||
546 .ValValProArgSerProSerAlaSerProArgSerSerSerLeuP 562

201 CTGCTCTCTATTATTGGCTCTTCTTGAAGCGCACACCGCGGCTGGG 152
   ||   :::::   ||   ::|||::|||   |||   |||
562 roArgThrSer.....SerSerProSerArgAlaGly 573

151 CGGCGTCTCCCGCGCGCCACTGGCCCTCGACGAACCTGCCGTC 102
   |||||   ::|||::|||::|||::|||   ::
574 ArgProHisHisaspGlnArgSerSerProHisLeuGlyArgSerLys 590

101 GACTCCACTGTCGATCTGGGTTTTCAGTTCGCGGAGATCC..... 59
   ::|||::|||::|||::|||::|||::|||::|||
590 sSerProProSerHisSerGlySerSerSerArgArgSerCysGlnG 607

58 .....GCTCGAAATTACTGCCTCTCTCCCGGAGGTAG 26
   ::|||::|||   |||   ::   ::
607 InGluHisCysLysProSerLysAsnGlyLeuLysGlySerGlySerLeu 623

25 CGGCATCGCTCTCATCT 8
   ::|||::|||::|||

```



624 HisHslsSerAlaSer 629

seq\_name: sp\_rodent:Q61359

seq\_documentation\_block:  
ID Q61359 PRELIMINARY; PRT; 243 AA.  
AC Q61359;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-DIC-2001 (TrEMBLrel. 19, Last annotation update)  
DE DYNAMIN (FRAGMENT).  
DNM OR BRADNM2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-SWISS WEBSTER; TISSUE-BRAIN;  
RA Stief A., van der Putten H.;  
RL Submitted (MAY-1994) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; L29457; AAA37319.1; -;  
DR MGD; MG1:107384; Dnm;  
DR InterPro; IPR003130; GED.  
DR Pfam; PF02212; GED; 1.  
DR SMART; SM00302; GED; 1.  
FT NON\_TER 1  
SQ SEQUENCE 243 AA; 26341 MW; 73BACE5E14EA1C9 CRC64;

alignment\_scores:  
Quality: 105.50 Length: 96  
Ratio: 2.198 Gaps: 4  
Percent Similarity: 50.000 Percent Identity: 35.417

alignment\_block:  
US-09-462-480-4/rev x Q61359 ..  
  
Align seg 1/1 to: Q61359 from: 1 to: 243

301 AGAGCCCATTTGG.....AGGACGCCCTGTCTGC 267  
||||||| :||| |||||||:::  
152 ArgSerProThrSerSerProThrProGlnArgArgAlaProAlaValPr 168  
||| |||||||:::  
266 TCCTCGTCGAGTTCGAGTATTGACGCCGCCCTGACGAATATTCGTCA 217  
||| |||||||:::  
168 oProAlaArgProGlySerArgGlyProAlaProGlyPro..... 181  
216 GATCTGTCGAGTTCCTCTCTCTCTATTCGCTCTTCCTTGGAAGCGCA 167  
|||||||:::  
182 .....ProAlaGlySer.....AlaLeuGlyGlyAla 191  
166 CCACGCCGCTGGCGGCCCTCCC GCCGCCGCCACTGCCCTGC 117  
||||||| ||:||:||:||:||:|||||:::  
192 ProProValProSerArgProGlyAlaSerProAspPropheGlyProPr 208  
:|||||||:::  
116 AACCAACCTGCGCTG.....ACTCCACCTGGTCGATCTGGGT 79  
:|||||||:::  
208 oProGlnValProSerArgProAsnArgAlaProProGlyValProSera 225  
78 TTTCAGTCGCGGAGATCCGCTCGAAATTAACCTGCCT 41  
|||||||:::  
225 fgsrGlyGlnAlaSerProSerArgProGluSerPro 237  
|||

seq\_name: sp\_virus:Q9DWG9

seq\_documentation\_block:  
ID Q9DWG9 PRELIMINARY; PRT; 252 AA.  
AC Q9DWG9;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE PR23.1.

GN R23.1.  
OS Rat cytomegalovirus (strain Maastricht).  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Betaherpesvirinae; Muromegalovirus.  
OX NCBI\_TaxID=79700;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-MAASTRICHT;  
RX MEDLINE=20366325; PubMed=10906222;  
R Vink C., Beuken E., Bruggeman C.A.;  
RT "Complete DNA sequence of the rat cytomegalovirus genome.";  
RJ J. Virol. 74:7656-7665(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-MAASTRICHT;  
RX MEDLINE=20473137; PubMed=11018281;  
RA Gruijthuisen Y.K., Beuken E., Bruggeman C.A., Vink C.;  
RT "Rat cytomegalovirus R89 is a highly conserved gene which expresses a  
spliced transcript";  
RL Virus Res. 69:119-130(2000).  
DR EMBL; AF232689; AAF99120.1; -;  
DR InterPro; IPR002965; P-rich\_extensions.  
DR PRINTS; PR01217; PRICHEXTENSN.  
SQ SEQUENCE 252 AA; 27216 MW; 42EE39D609F4C0E2 CRC64;

alignment\_scores:  
Quality: 104.00 Length: 112  
Ratio: 2.122 Gaps: 5  
Percent Similarity: 43.750 Percent Identity: 33.036

alignment\_block:  
US-09-462-480-4/rev x Q9DWG9 ..  
  
Align seg 1/1 to: Q9DWG9 from: 1 to: 252

277 CCTGCTGCTGCTCTCTCTGCGCCCTCGAGTATTGGACGCCGCTGACGA 228  
|||||||:::||| |||||:::  
124 ProAlaSerSerProValArgAspProProGlyArgThrProArgPr 140  
227 ATATTCGTCGAGATCTCGTCGAGTTCCTGCTTCTGCTTATTGGCTGCTTC 178  
|||||||:::||| |||||:::  
140 oThrProCysArgAlaProArgValProTyfLeuCys..... 152  
177 TTGGAAGCGCACACCACGCCGCTGGCGCGCTCCCCG..... 140  
153 .....ProArgProGlyArgProArgAlaAsnMetPheGly 164  
139 .....CCGCGCGCGCCACTGGCCCTCGAACGAACCTGCGCTC 102  
|||||||:::||| |||||:::  
165 LeuSerArgArgGlyArgGlyThrGlyAlaAlaSerArgAlaProSe 181  
101 GACT.....CCACTGCTC 88  
|||||||:::||| |||||:::  
181 rAlaAspAlaProProArgProArgAlaAlaAraProAspProArgGlyA 198  
87 GATCTGGGTTTTTCAGGTCGCGGAGATCCGCTCGAAAATACTCCGCTCCT 38  
|||||||:::||| |||||:::  
198 rgThr.....SerGlySerArgArgSerGlyThrProArgCysPro... 211  
37 GCCCGAGGGTAGCGGCATCGGCTCTTCATCTCTGCCA 2  
||| |||||:::||| |||||:::  
212 .....ArgSerArgAlaAlaSerArgPro 219

seq\_name: sp\_rodent:Q9CUW3

seq\_documentation\_block:  
ID Q9CUW3 PRELIMINARY; PRT; 255 AA.  
AC Q9CUW3;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE 2900057H02RIK PROTEIN (FRAGMENT).

GN 2900057H02RIK.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=HIPPOCAMPUS;  
RX MEDLINE=21085660; PubMed=11217851;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
RA Alizadeh K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
RA Hayashizaki Y.,  
RT "Functional annotation of a full-length mouse cDNA collection."  
RL Nature 409:685-690(2001).  
DR EMBL; AK013719; BAB28968.1; ..  
DR MGD; MGI:1924225; 2900057H02RIK.  
DR InterPro; IPR002965; P-rich\_extensn.  
DR PRINTS; PR01217; PRICHEXTENS.  
FT NON\_TER 1  
SQ SEQUENCE 255 AA; 27606 MW; 07F5DE6B902D5FE6 CRC64;

alignment\_scores:  
Quality: 103.50 Length: 116  
Ratio: 1.754 Gaps: 5  
Percent Similarity: 50.862 Percent Identity: 31.897

alignment\_block:

US-09-462-480-4/rev x Q9CUW3

Align seg 1/1 to: Q9CUW3 from: 1 to: 255

283 ACACGGCT.....GCTGCTGCTCTCGTCGGCCCTCGAGTATTGGAGC 240  
|||||  
40 ThrAlaProGlyGlnAlaAlaGlyProGlyArgProGlnGlnProArgG1 56  
239 C.....CGGCCTGACGAATATTCGTGCA 217  
56 nGlyAspProGlyProAlaGlyProHisArgProArgAlaPro\*\*GlnA 73  
216 GATCTGCTCGAGTTCCTGCTCTGCTATTGCTGCTCTTCTTGAAGCGCA 167  
|||||  
73 rGluGlyArgAlaProGlyPro.....GluAla 82  
166 CCACCGCGGCTGGCGGCGCTGCCCG....CCGCGCGCGCCACTGGCCC 120  
|||||  
83 ProAlaGlyProGlyGlnProProProProProProProProProProGly 99  
119 TGCAACGAACCTGCGCTGACT..... 98  
99 uSerThrArgArgProThrProProProProProProProProProProPro 116  
97 .....CCACCTGGTGATCTGGGTTTTCAGGTTCGCGCGAGA 62  
116 erGly\*\*GlyArgProProGlyArgLysGluGlyThrArgAlaArg 132  
|||||  
61 TCCGCTGGAATACCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 14  
::: |||||

133 AspProArgLysProArgArgAlaLysSerSerArgArgSer 148  
seq\_name: sp\_bacteria:Q93R69

seq\_documentation\_block:

ID Q93R69 PRELIMINARY; PRT; 620 AA.

AC Q93R69;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE LEVANASE.  
OS Microbacterium laevaniformans.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Micrococciaceae; Microbacteriaceae; Microbacterium.  
OX NCBI\_TaxID=36807;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 15953;  
RA Song E., Cha J.;  
RT "Microbacterium laevaniformans levanase-complete cds."  
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB062062; BAB59060.1; ..  
SQ SEQUENCE 620 AA; 67667 MW; 82B4909328CBAAC2 CRC64;

alignment\_scores:

Quality: 102.00 Length: 105  
Ratio: 1.821 Gaps: 5  
Percent Similarity: 53.333 Percent Identity: 34.286

alignment\_block:

US-09-462-480-4 x Q93R69

Align seg 1/1 to: Q93R69 from: 1 to: 620

29 CCCTGGGCGAGGAGG.....CAGGTATTTTCAGCGGATCTCCGGCGAC 72  
|||||  
520 ProSerGlyAspGlnGlyLeuAlaLeuThrSerGlyGlyProAlaVa 536  
73 CTGAAA.....CCCCA 83  
536 lPheArgAspLeuArgIleThrGluPheAlaAspLeuAlaHisValProA 553  
84 GA...TCGACCAAGTTCGAGTCGACGCGAGGTCGTTGTCAGCGCGCGAGTGC 130  
|||  
553 rGAlaSerSerArgThrSer.....ArgAlaThrArg 563  
131 GCGCGCGCGGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180  
:::|||||  
564 GlyAlaArgGlyGlyArgProProAlaProSerProGlySerValArgAr 580  
181 GCACCCAATAGCAGACGAGGAACTCG.....ACGAGATCTC 218  
::: |||  
580 gArgAlaIleCysAlaGlyArgArgAlaArgAlaTrpProThrArgSerP 597  
219 GACGAATATTCGTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 268  
|||||  
597 rOGlyArgThrProArgProAlaSerSerArgArgArgProSerArgSer 613  
269 AGCAGCAGCGCGCTGT 283  
614 ThrGlyThrArgCys 618

seq\_name: sp\_rodent:Q9JKA7

seq\_documentation\_block:

ID Q9JKA7 PRELIMINARY; PRT; 1198 AA.

AC Q9JKA7;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE HYPERPOLARIZATION-ACTIVATED, CYCLIC NUCLEOTIDE-GATED POTASSIUM CHANNEL  
DE 4.

GN HCN4.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=BRAIN;  
RA Monteggia L.M., Eisch A.J., Tang M.D., Kaczmarek L.K., Nestler E.J.;  
RT "Cloning and Localization of the Hyperpolarization-Activated Cyclic  
Nucleotide-Gated Channel Family in Rat Brain."  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF247453; AAF62176.1;  
DR InterPro; IPR000636; Cation\_chan\_non\_lig.  
DR InterPro; IPR000595; cAMP\_binding.  
DR Pfam; PF00027; cAMP\_binding; 1.  
DR Pfam; PF00520; ion\_trans; 1.  
DR SMART; SM00100; cAMP; 1.  
DR PROSITE; PS00888; CNMP\_BINDING\_1; UNKNOWN\_1.  
DR PROSITE; PS00442; CNMP\_BINDING\_3; 1.  
KW Ionic channel.  
SQ SEQUENCE 1198 AA; 128760 MW; 6B92B8F9452F760F CRC64;

alignment\_scores:  
Quality: 102.00 Length: 133  
Ratio: 2.125 Gaps: 5  
Percent Similarity: 36.090 Percent Identity: 28.571

alignment\_block:

US-09-462-480-4/rev x Q9JKA7 ..

Align seg 1/1 to: Q9JKA7 from: 1 to: 1198

301 AGAAGCCATTTTCGGAGGACAGCGCTGCTGCTCCT..... 263  
|||||  
933 ArgSerProGlnAlaLagInProProProProLeuProGlyAlaArgG1 949  
262 .....CGTC 259  
949 YGlyLeuGlyLeuLeuGluHisPheLeuProProProSerSerArgS 966  
258 GGCCTCGAGT.....ATTGGACGGCGCTGACGAATATTCG 221  
|||||  
966 ArgProSerSerProGlyGlnLeuGlyGlnProProGlyGluLeuSer 982  
220 TCGAGATCTCGTCGAGTCTCTGCTTCTGCTATTGCTGCTTCTGGAAG 171  
983 ProGlyLeuAlaAlaGlyProProSerThr.....ProG1 994  
170 CGCACCCACCGCGCTGGCGGCGCG..... 146  
994 uThrProArgProGluArgProSerPheMetAlaGlyAlaSerGlyG 1011  
145 .....TCCCGCGCGCGC 134  
1011 LysAlaSerProValAlaPheThrProArgGlyGlyLeuSerProGly 1027  
133 CGCGCCACTGCGCTGCAACGACCTGCGCTGACCTCCACCTGGTCGATC 84  
1028 HisSerProGlyProArgThrPheProSerAlaProProArgAlaSe 1044  
83 TGGGTTTTCAGGTGCGCGGAGATCCGCTCGAAATTAACCTGCTCCTGCC 35  
1044 rGlySerHisGlySerLeu.....LeuLeuProProAla 1055  
|||

seq\_name: sp\_rodent:070495

seq\_documentation\_block:

ID 070495 PRELIMINARY; PRT; 897 AA.  
AC 070495;  
DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE PLENTY-OF-PROLINES-101.  
GN SRRM1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=BRAIN;  
RA Vayssières B.M., Camonis J.H.;  
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF062655; AAC17422.1;  
DR MGD; MGI:1858303; Strim1.  
DR InterPro; IPR002483; PWI.  
DR Pfam; PF01480; PWI; 1.  
DR SMART; SM00311; PWI; 1.  
SQ SEQUENCE 897 AA; 101166 MW; 62160BEB6772BD10 CRC64;

alignment\_scores:  
Quality: 101.00 Length: 114  
Ratio: 1.870 Gaps: 6  
Percent Similarity: 47.368 Percent Identity: 34.211

alignment\_block:

US-09-462-480-4/rev x 070495 ..

Align seg 1/1 to: 070495 from: 1 to: 897

298 AGCCCATTTTCGAGGACAGCGCTGCTGCTGCTGCTGCGCCCTCGAG 249  
|||||  
522 SerProSerArgSerAlaSerProSerProArgLysArgGlnLysGluTh 538  
248 T.....ATTGGACCGCG.....CCTGACCAATATT 223  
538 rSerProArgMetGlnMetGlyLysArgTrpGlnSerProValThrLysS 555  
222 CGTCGAGATCTCGTCGAGTTCCTGCTTCTGCTTATTGCTGCTTCTTGA 173  
555 erSerArgArgArgArgSerProProProProAlaArgArgArgArg 571  
172 AGC.....GCACACCGCGCGCTGGCGCGCTGCCCGCGCGCGCG 132  
572 SerProSerProAlaProProProProProProProProProProArg 588  
131 CGCCACTGCGCCCTCCA.....ACGACCTGCGCTGCGACTCCAC 94  
588 gArgArgSerProThrProProProProProProProProProProSer 605  
93 CTGGTCGATCTGGGTTTTCAGGTCGCGGAGATCCGCTCGAAATTAACCTG 44  
605 roArgArg.....ArgSerProSerProArgArgTyrSer 616  
43 CTCTCTGCGCGAGGTAGCGCATCGGCTTTCATCTCTGCGCA 2  
617 ProProIleGln.....ArgArgTyrSerProSerProPro 628

seq\_name: sp\_bacteria:087852

seq\_documentation\_block:

ID 087852 PRELIMINARY; PRT; 859 AA.  
AC 087852;  
DT 01-NOV-1998 (TREMBLrel. 08, Created)  
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE PUTATIVE BETA-GLUCOSIDASE.  
GN SC8A6.18.  
OS Streptomyces coelicolor.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1902;  
RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Seeger K.J., Harris D.;  
 RL Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Parkhill J., Barrell B.G., Rajandream M.A.;  
 RL Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA MEDLINE=97000351; PubMed=8843436;  
 RX Redenbach M., Kleser H.M., Denapante D., Eichner A., Cullum J.,  
 RA Kinashi H., Hopwood D.A.;  
 RT "A set of ordered cosmids and a detailed genetic and physical map for  
 the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
 RL Mol. Microbiol. 21:77-96(1996).  
 DR EMBL: AL031013; CAA19790.1;  
 DR InterPro: IPR001764; Glyco\_hydro\_3;  
 DR InterPro: IPR002772; Glyco\_hydro\_3C;  
 DR Pfam: PF00933; Glyco\_hydro\_3; 1.  
 DR Pfam: PF01915; Glyco\_hydro\_3\_C; 1.  
 DR PRINTS: PR00133; GLHIDRLASE3.  
 SQ SEQUENCE 859 AA; 89455 MW; BD7EAFEAE4FD59F CRC64;

alignment\_scores:  
 Quality: 99.50 Length: 92  
 Ratio: 2.117 Gaps: 5  
 Percent Similarity: 51.087 Percent Identity: 38.043

alignment\_block:  
 US-09-462-480-4 x 087852

Align seg 1/1 to: 087852 from: 1 to: 859

39 GGAGGAGGTAATTCGAGCGGATCTCCGGGACCTGAAACCCAGATCG 88  
 287 GlyGlyArg.....AspValSerArgSerProValProGlyArg 299  
 89 ACCAGGT.....GGAGTCAGCGCGAGTTCGTTGTCAG 120  
 299 gProAlaAspSerProGlyAlaGluGlyAlaAspGlyGlyAlaGlyAlaG 316  
 121 GCCAGTGGCGCGCGCGGGGAGCGCGCGAGCGCGGT.....164  
 316 hProSerSer.....GlyAlaGluGlyLeuProGlyArgGlyProAla 330  
 165 ...GGTGGCTTCCAGAGCAGCGCAATAGCAGAGCAGCACTCGACG 211  
 331 HisGlyAlaLysProSerGlyProArgProArgAlaGlyAspGlyArg 347  
 212 AGATCTGACGAATATTCGTACGCGCGCGTCCAACTACTGAGGCGCGAC 261  
 347 gAlaLeu.....AlaArgA 352  
 262 GAGGAGCAGCAGCGCGCTGCTCTC 287  
 352 rgAlaValAlaAlaGlyAlaValLeu 360

seq\_name: sp\_bacteriap:Q9RX36

seq\_documentation\_block:

ID Q9RX36 PRELIMINARY; PRT; 873 AA.  
 AC Q9RX36;  
 DT 01-MAY-2000 (TRENBLrel. 13, Created)  
 DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)  
 DE 01-JUN-2001 (TRENBLrel. 17, Last annotation update)  
 GN PENICILLIN-BINDING PROTEIN 1.  
 OS Deinococcus radiodurans  
 OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.

OX NCBI\_TaxID=1299;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=R1;  
 RX MEDLINE=20036896; PubMed=10567266;  
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,  
 RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,  
 RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,  
 RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,  
 RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,  
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,  
 RA Fraser C.M.;  
 RT "Genome sequence of the radioresistant bacterium Deinococcus  
 radiodurans R1.";  
 RL Science 286:1571-1577(1999).  
 DR EMBL: AE001907; AAF10059.1;  
 DR TIGR: DR0479;  
 DR InterPro: IPR002965; P-rich\_extensn.  
 DR InterPro: IPR001264; Transglycosyl.  
 DR InterPro: IPR001460; Transpeptidase.  
 DR Pfam: PF00912; Transglycosyl; 1.  
 DR Pfam: PF00905; Transpeptidase; 1.  
 DR PRINTS: PR01217; PRICHEXTENS.  
 DR ProDom: PD001895; Transglycosyl; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 873 AA; 95571 MW; 8B859A6B9919E086 CRC64;

alignment\_scores:

Quality: 99.50 Length: 135  
 Ratio: 1.746 Gaps: 9  
 Percent Similarity: 42.222 Percent Identity: 32.593

alignment\_block:

US-09-462-480-4/rev x Q9RX36

Align seg 1/1 to: Q9RX36 from: 1 to: 873

298 AGCCCATTTGCGAGGACACGCGCTGCTGCTGCTGCTGCGCCCTCGAG 249  
 724 ThrProAlaValThrAlaPro.....GluProProGlnProAlaTh 738  
 248 T.....ATTGAGCGCGGCTGACGAATATTCGTGCGAGATCTGTC 208  
 738 rThrThrProAlaGlyArgSerProAlaProLeuArgArgArgThrArg 755  
 207 GAGTT.....CCTGCTTCTGCTTATTGGCTGCTCTT 176  
 755 rgThrLeuProArgGlnThrSerProArgArgAla.....Leu 767  
 175 GGAAGCGCACCCAGCGCGCTGGCGGCGG...TCCCGCGCG... 137  
 768 ProSerArgThrCysArgProArgArgProLeuSerProProArgArg 784  
 136 .....CGCGCGCCCACTGGCCCT. 119  
 784 gProArgProGlyAlaThrArgArgAlaAsnArgArgArgThrSerProp 801  
 118 .....GCAACGAACCTCCGTCG..... 101  
 801 roSerAlaThrArgCysProThrCysProArgArgArgLeuSerArgCys 817  
 100 .....ACTCACCTGGTCG 87  
 818 ProAlaThrProArgProLeuSerArgArgProValThrProProGlyArg 834  
 86 ATCTGGGTTTTCAGTCGCGGAGATCCGCTCGAAATTAACCTGCTCCTG 37  
 834 gPro...PheGlnAlaArgArgGlyProArgLeuArgGlyProProt 850  
 36 CCGGA 32  
 850 hrArg 851

seq\_name: sp\_vertibrate:057580

```
seq_documentation_block:
ID   O57580      PRELIMINARY;          PRT;   1151 AA.
AC   O57580;
DT   01-JUN-1998 (TREMBLrel. 06, Created)
DT   01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DE   HIGH MOLECULAR MASS NUCLEAR ANTIGEN (FRAGMENT).
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98034404; PubMed=9365273;
RA   Shimada K., Harata M., Mizuno S.;
RT   "A nuclear matrix-associated high molecular mass nuclear antigen,
RT   HUNA, of chicken and marked decrease of its immunoreactivity during
RT   the progression of S phase."
RL   J. Cell Sci. 110:3031-3041(1997).
DR   EMBL; D88440; BAA24137.1; -.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   InterPro; IPR001778; POA_allergen.
DR   PRINTS; PR00015; GPOSANCHOR.
DR   PRINTS; PR00833; POAALLERGEN.
FT   NON_TER 1
FT   NON_TER 1151
FT   SEQUENCE 1151 AA; 109708 MW; 2403F6835F9A2AB3 CRC64;
```

```
alignment_scores:
Quality: 99.00      Length: 115
Ratio: 2.062       Gaps: 4
Percent Similarity: 41.739   Percent Identity: 26.957

alignment_block:
US-09-462-480-4/rev x O57580 ..
Align seg 1/1 to: O57580 from: 1 to: 1151
```

```
298 AGCCATTGGAGGACAGCCGCTGCTCTGCTGGCCCTCGAG 249
      :::::  |||||  |||||  |||||  |||||
16 AlaProSerProAlaProAlaProAlaProAlaProAlaPro..... 30

248 TATTGGAGCGGCGCTGACGAATATTCGTCGAGATCTCGTAGATTCTGT 199
      |||  ::  |||||
31 .....ProArgProLysTrpValProI 38

198 CTTCTGCTTAT.....TGGCTGCTT 179
      :::::  |||||
38 leAlaGluLeuHisProAlaAlaProGlnProProLysTrpValPro 54

178 CTTGGAACGCGACACCGCGCTGGG...CGGCGCTCCCGCGCGCG 132
      :::::  |||||  |||||  |||||  |||||  ::
55 ileGlyGlyAlaProProProGlyThrGluProThrProProSerLy 71

131 C.....GCCACTG 124
      :
71 sProThrAspGlyAlaAspAlaAlaProLysAlaSerAlaGluLeuThrs 88

123 GCCCTGACCAACCTGCCCTGACTCCACTGGTCGATCTGGGTTTCA 74
      ::|||  |||||  |||  :::::  |||
88 erProProAlaSerProProAspGlyProLysAlaProSer 104

73 GGTGCGCGGAGATCCGCTCGAAATACCTGCTCTGCTGCCGAGGG 29
      |||  :::::  ::  |||||  :::::  |||
105 GlyAlaGlyGluAlaGluAlaGlyThrProProSerGingly 119
```

seq\_name: sp\_plant:09XJ18

seq\_documentation\_block:

```
ID   Q9XJ18      PRELIMINARY;          PRT;   317 AA.
AC   Q9XJ18;
DT   01-NOV-1999 (TREMBLrel. 12, Created)
DT   01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DE   SIMILAR TO HUMAN MRNA FOR ALPHA 1.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzaceae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. NIPPONBARE;
RA   Sasaki T., Nagamura Y., Yamamoto K.;
RT   "Oryza sativa nipponbare(CA3) genomic DNA, chromosome 6, PAC
RT   clone:P0680A03."
RL   Submitted (FEB-1999) to the EMBL/GenBank/DBDJ databases.
DR   EMBL; AB023482; BAA78734.1; -.
SQ   SEQUENCE 317 AA; 34259 MW; 060873B38658BB0A CRC64;
```

```
alignment_scores:
Quality: 98.50      Length: 144
Ratio: 1.698       Gaps: 8
Percent Similarity: 40.278   Percent Identity: 31.944

alignment_block:
US-09-462-480-4/rev x Q9XJ18 ..
Align seg 1/1 to: Q9XJ18 from: 1 to: 317
```

```
280 GCGCTGCTGCTCTCTCTCGCGCCCTCGAGTATTGACGCCGCCCTGA 231
      |||||  ||:::  |||||  :::::  |||  ::
48 AlaPro...AlaGlyGluArgArgGlyAlaAlaAspGlyGlyTrpGluG 63

230 CGAATATTTCGAGATCTCGTCTCGAGTTCTCTCTCTTAT..... 188
      :|||  ||:::  |||||  :::::  ::
63 uGluArgSerAlaArgTyrArgArgIleMetArgGlyArgAsnGlyHis 80

187 .....TGGCTGCTTCTGGAGCGCACCCGCGCCCTGGGGG..... 149
      |||||  |||||  |||||  ::|||
80 yrProTrpLeuLeu.....ProProThrGlyAlaAlaArgAlaPro 93

149 ..... 149

94 LeuLeuSerArgArgSerProLeuMetSerProLeuProLeuGlySerTh 110

148 ....CCGTCGCCCGCGCGCGC.....GCCA 127
      |||  |||||  |||||  ::
110 rThrProLeuProProArgArgGlyGlnArgLeuArgArgHisSerSerH 127

126 CTGCCCCCTGCAACGAACCTGCCGTCGACTCCACCT..... 92
      |||||  |||||  |||||  |||||
127 isProProAlaProLeuHisProSerThrProProGlyHisArgGlySer 143

91 .....GGTCGATCTGGGTTTTCAGTCCGCGGAGATCCGCTCGAAA 51
      |||||  |||  ::|||  ::
144 GlySerArgGlyArgGlyCysGlyArgLeuLeuAlaProLeuPr 160

50 TTACTGCTCTCT.....GCCCGAGGTAGCGGC 22
      |||||  |||  |||  |||
160 oLeuLeuProProThrLysThrThrThrMetAlaThrGlySerArgA 177

21 ATCGS.....TCTTCATCTCTGCCA 2
      :|||  |||||  |||
177 rgArgArgThrArgThrSerSerSerGlnPro 187
```

seq\_name: sp\_plant:Q94EV7

```
seq_documentation_block:
ID   Q94EV7      PRELIMINARY;          PRT;   406 AA.
```

AC Q94EV7;  
DT 01-DEC-2001 (TReMBLrel. 19, Created)  
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
DE MAP KINASE KINASE.  
GN MAPK1.  
OS Zea mays (Maize).  
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;  
OC Panicoideae; Andropogoneae; Zea.  
OX NCBI|TaxID=4577;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TRANSPOSOME-HUCK1;  
RA Fu H., Zheng Z., Dooner H.K.;  
RT "Large differences in recombination rates within adjacent gene-dense  
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
KW EMBL; AF391808; AAK73104.1; -;  
KW Kinase.  
SQ SEQUENCE 406 AA; 44664 MW; DBFIDE6568F47DF0 CRC64;

alignment\_scores:  
Quality: 98.50 Length: 100  
Ratio: 1.894 Gaps: 5  
Percent Similarity: 52.000 Percent Identity: 38.000

alignment\_block:

US-09-462-480-4/rev x Q94EV7

Align seg 1/1 to: Q94EV7 from: 1 to: 406

286 AGGACAGCGCTGCTGCT...GCTCCTCGTCGCGCTCGAGTATTGGACG 240  
|||||:|||||:||||| ||| |||||||||: |||  
56 ArgThrThrProCysAlaArgAlaAlaSerArgProSerProAlaAr 72  
239 CCGGCTGACGAATTCGTGCGA.....T 214  
||||| :|: |||||  
72 gArgAlaAlaArgAlaAlaAlaArgIleSerProSerProCysProSerA 89  
213 CTCGTCGAGTTCCTGCTCTGCTTATTGGCTGCTTCTTGGAGCGCACCA 164  
:||||| ||| ||| ||| ||| :|||  
89 IaArgSerProProProSerProSerArgSerLeuSerArgGlnProPro 105  
163 dCGCGGCTGGCGCGCGCTGCCCGCGCGCGCCACTGCCCTGCAAC 114  
||| ||||| :||||| |||||: |||: |||  
106 ProProProGly.....GlyProProArgAlaSer..... 116  
113 GAACCTGCGCTGACTCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 64  
||| ||||| ||||| |||||: |||: |||  
117 .....ProArgProProProProArgSerSerSerSerSerSerProp 131  
63 GATCGCTCGAATACCTGCTCCTCGCCGAGGTACCGCATCGCT 16  
||| |||||: |||: ||||| |||  
131 roLeuAla.GluLeuGluArgValArgArgValGlySerGlyAlaGly 146

seq\_name: sp\_human:Q9UQ39

seq\_documentation\_block:

ID Q9UQ39 PRELIMINARY; PRT; 956 AA.

AC Q9UQ39;  
DT 01-MAY-2000 (TReMBLrel. 13, Created)  
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
DE RNA BINDING PROTEIN (FRAGMENT).

OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI|TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Ohtaki S., Umeki K., Sawada Y.;

RT "Homo sapiens mRNA for RNA binding protein, partial cds.";  
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB016088; BAA83714.1; -;  
DR InterPro; IPR002965; P\_Rich\_extensn.  
DR PRINTS; PR01217; PRICHEXTENS1.  
FT NON\_TER 1  
FT NON\_TER 956  
SQ SEQUENCE 956 AA; 106462 MW; C9FEAD07B49F5011 CRC64;

alignment\_scores:

Quality: 98.50 Length: 140  
Ratio: 1.669 Gaps: 5  
Percent Similarity: 42.143 Percent Identity: 27.143

alignment\_block:

US-09-462-480-4 x Q9UQ39

Align seg 1/1 to: Q9UQ39 from: 1 to: 956

14 AGACGATGCGCTACCTCGCGGAGGAGGAGGTAATTTTCG..... 55  
:||||| ||| |||:|||||:|||||: |||  
407 LysProAlaProAlaProGlySerHisArgGluIleSerSerProTh 423  
55 ..... 55  
423 rSerLysAsnArgSerHisGlyArgAlaLysArgAspLysSerHisSerH 440  
56 .....ACGGATCTCCGCGGACCTGAAAA 79  
||| ||| ||||| ||| |||  
440 lThrProSerArgArgMetGlyArgSerArgSerProAlaThrAlaLys 456  
80 .....CC 81  
457 ArgGlyArgSerArgSerArgThrProThrLysArgGlyHisSerArgSe 473  
82 CAGATCGACACAGTGGAGTCCA.....CGCAGGTTCGTTGCAGGG 122  
||||| :||| ||| |||  
473 rArgSerProGlnTrpArgArgSerArgSerAlaGlnArgTrpGlyArgS 490  
123 CCAGTGGCGCGCGCGGCGGACGG.....CGCCCGAGCGCGCGG 163  
:|: ||| ||||| ||| ||| |||  
490 erArgSerProGlnArgArgGlyArgSerArgSerProGlnArgProGly 506  
164 TGGTGGCGCTTCCAAAGAG...CAGCCAATAAGCAGAGCAAGCACTCGAC 210  
||| |||: ||| :||||| |||: |||  
507 TrpSerArgSerArgAsnThrGlnArgArgGlyArgSerArgSerAlaAr 523  
211 GAGATCTCGACGAATATTCGTCAGCGCGGCTCCAAATCTCGAGGCGCGA 260  
|||: ||| |||||: |||: |||: |||  
523 gArgGlyArgSerHisSerArgSerProAlaThrArgGlyArgSerArgS 540  
261 CGAGGAGCAGCAGCGCGC 280  
540 erArgThrProAlaArgArg 546

OM of: US-09-462-480-3 to: A\_Geneseq.032802:\* out\_format : pfs  
Date: Jul 22, 2002 1:22 AM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:  
-MODEL=frame.n2p.model -DEV=xih  
-O=/cgn2.1/USPTO.spool/US09462480/runat\_18072002.164418.19413/app\_query.fasta.1.2850  
-DB-A\_Geneseq.032802 -QFMT=fastan -SUFFIX=rag -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-XGAPOP=4.500 -XGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsum62  
-TRANS=human40 cdi -LISF=45 -DOCALIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09462480.@CGN1.1.57 -NCPU=6 -TCPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

Search information block:  
Query: US-09-462-480-3  
Query length: 481  
Database: A\_Geneseq.032802:\*  
Database sequences: 747574  
Database length: 11107396  
Search time (sec): 134.850000

score\_list:

Sequence	Strd Orig	ZScore	EScore	Len	! Documentation
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1997.DAT:AAW32452 +	...	713.00	1059.30	2.0e-51	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1997.DAT:AAW32384 +		713.00	1059.30	2.0e-51	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1998.DAT:AAW61705 +		713.00	1059.30	2.0e-51	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1998.DAT:AAW64338 +		713.00	1059.30	2.0e-51	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1999.DAT:AAW39135 +		713.00	1059.30	2.0e-51	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1999.DAT:AAW38992 +		713.00	1059.30	2.0e-51	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2001.DAT:AAW19844 +		713.00	1059.30	2.0e-51	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1998.DAT:AAW72329 +		713.00	1059.23	2.0e-51	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1999.DAT:AAW21946 +		713.00	1059.23	2.0e-51	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1998.DAT:AAW72928 +		373.00	559.56	3.7e-23	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1999.DAT:AAW21945 +		373.00	559.56	3.7e-23	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2000.DAT:AAW54248 +		147.50	215.51	0.0002	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2000.DAT:AAW54248 +		145.50	216.42	0.0003	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2001.DAT:AAW79841 +		141.50	195.79	0.0009	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1999.DAT:AAW24091 +		140.00	200.22	0.0020	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2001.DAT:ABG21932 +		136.50	195.59	0.0031	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1998.DAT:AAW79128 +		134.00	196.80	0.0031	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1997.DAT:AAW31855 +		134.00	190.16	0.0034	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1997.DAT:AAW31852 +		134.00	187.68	0.0035	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2001.DAT:AAW31370 +		133.50	198.23	0.0033	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1998.DAT:AAW20852 +		132.50	196.94	0.0040	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2001.DAT:AAW61738 +		132.50	185.23	0.0046	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2001.DAT:AAW39141 +		131.00	183.85	0.0061	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1998.DAT:AAW72204 +		128.50	167.31	0.0116	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2001.DAT:ABG21919 +		127.50	185.32	0.0110	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2001.DAT:ABG30150 +		127.50	183.22	0.0113	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2001.DAT:AAW31516 +		127.50	169.28	0.0134	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2000.DAT:AAW12000 +		127.50	169.26	0.0134	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1997.DAT:AAW18664 +		125.50	180.85	0.0164	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2001.DAT:ABG03626 +		125.00	184.17	0.0172	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2001.DAT:ABW70063 +		124.50	178.11	0.0202	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2001.DAT:ABW11016 +		124.00	175.68	0.0227	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA18528 +		123.00	180.33	0.0255	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2001.DAT:ABW65293 +		122.50	166.64	0.0329	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2001.DAT:ABG04558 +		121.00	176.96	0.0376	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2001.DAT:ABG19764 +		121.00	175.80	0.0381	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1998.DAT:AAW68408 +		121.00	173.73	0.0391	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1992.DAT:AAW28150 +		121.00	175.02	0.0395	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2000.DAT:AAW28595 +		121.00	169.39	0.0413	
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2001.DAT:ABG20365 +		121.00	165.11	0.0435	

/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1999.DAT:AAW13464 - 120.50 163.25 0.04  
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1999.DAT:AAW76734 - 120.00 162.05 0.05  
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2000.DAT:AAW17737 - 119.50 175.43 0.05  
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2000.DAT:AAW17736 - 119.50 174.53 0.05  
/SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2000.DAT:AAW17735 - 119.50 171.28 0.05

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1997.DAT:AAW32452

seq\_documentation\_block:  
ID AAW32452 standard; Protein; 368 AA.

XX AC AAW32452;

XX DT 09-JAN-1998 (first entry)

XX DE Mycobacterium tuberculosis antigen Tb37-Fl.

XX KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;  
XX KW skin testing; M.tuberculosis.

XX OS Mycobacterium tuberculosis.

XX FH Key Location/Qualifiers  
XX FT Misc-difference 10 /note= "Any amino acid"

XX PN WO9709428-A2.

XX PD 13-MAR-1997.

XX PF 30-AUG-1996; 96WO-US14674.

XX PR 12-JUL-1996; 96US-0680574.

XX PR 01-SEP-1995; 95US-0523436.

XX PR 22-SEP-1995; 95US-0533634.

XX PR 22-MAR-1996; 96US-0620874.

XX PR 05-JUN-1996; 96US-0659683.

XX PA (CORI-) CORIXA CORP.

XX PI Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;

XX PI Twardzik DR, Vedvick TH;

XX DR WPI; 1997-192903/17.

XX PT New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are  
XX PT useful in vaccines for prevention or treatment of tuberculosis, also  
XX PT for diagnosis

XX PS Example 3; Page 146-147; 168pp; English.

XX CC A new immunogenic polypeptide has been developed comprising an  
XX CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or  
XX CC its variant differing only in conservative substitutions and/or  
XX CC modifications). The present sequence represents a M.tuberculosis  
XX CC antigen, Tb37-FL. The immunogenic protein, and fusion proteins  
XX CC containing one or more of the proteins or one of the proteins plus  
XX CC ESAN-6, are useful in vaccines preferably when formulated with a  
XX CC non-specific adjuvant, to induce an immune response against  
XX CC M.tuberculosis (for treatment or prevention).

XX SQ Sequence 368 AA;

alignment\_scores:  
Quality: 713.00 Length: 143  
Ratio: 5.057 Gaps: 0  
Percent Similarity: 98.601 Percent Identity: 96.503

alignment\_block:  
US-09-462-480-3 x AAW32452 ..  
Align seg 1/1 to: AAW32452 from: 1 to: 368

```
1 CTGCAGCAGGTGACGTCGTTGTTTTCAGCCAGGTGGCGCGCACCGCGCGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGlyG1 242
51 CAACCCAGCCAGCAGGAGGAGCGGACATGGGCTGCTCGGCACCGATC 100
|||||
242 YAsnProAlaAspGluGluAlaGlnMetGlyLeuLeuGlyThrSerP 259
101 CGCTGTGCAACATCCGCTGCTGCTGATCAGCCAGCGCCGCGGCGG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275
151 GGCCTGCTGCGCGGAGTGCCTACCTGCGCAGGTGGCTGCTTACCCG 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292
201 CACGCCGCTGATGCTCAGCTGATCGAAGAACCGGTTGCCCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309
251 TGCCGGCGGCTGTTGCGGATCGTCGGTGACGGTGGCGCGCTCCGGTG 300
|||||
309 etProAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
301 GTTCCGGGAGCAGTGGCGCAGGTTGCGCAATCCGCGCGCTCCACGACC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
351 GGGTCTGTCGCGCGGCGCACGCTCGCGCAGGAGCGGTGAAGAAGACG 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAspG 359
401 AGGACGACTGGGACGAGGACGACTGG 429
359 luAspAspTrpAspGluAspAspTrp 368
```

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:AAW32384

seq\_documentation\_block:

ID AAW32384 standard; Protein; 368 AA.

AC AAW32384;

DT 13-JAN-1998 (first entry)

DE Mycobacterium tuberculosis antigen Tb37-FL.

KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;

KW skin testing; M.tuberculosis.

XX Mycobacterium tuberculosis.

FH Key Location/Qualifiers

FT Misc-difference 10 /note= "Any amino acid"

FT WO9709429-A2.

XX PD 13-MAR-1997.

XX PF 30-AUG-1996; 96WO-US14675.

XX PR 12-JUL-1996; 96US-0680573.

PR 01-SEP-1995; 95US-0523435.

PR 22-SEP-1995; 95US-0532136.

PR 02-MAR-1996; 96US-0620280.

PR 05-JUN-1996; 96US-0658800.

XX (CORI-) CORIXA CORP.

XX Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;

PI Twardzik DR, Vedvick TH;

XX

DR WPI; 1997-192904/17.

XX New immunogenic polypeptide(s) from soluble M. tuberculosis antigens

PT - useful for diagnosis of M. tuberculosis infection

XX Example 3; Page 159-161; 190pp; English.

XX A new immunogenic polypeptide has been developed comprising an  
CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or  
CC its variant differing only in conservative substitutions and/or  
CC modifications). The present sequence represents a M.tuberculosis  
CC antigen, Tb37-FL. The immunogenic polypeptide can be used to diagnose  
CC M.tuberculosis infection by forming complexes with specific  
CC antibodies in the sample. Fragments of DNA encoding the immunogenic  
CC polypeptide can be used as diagnostic primers or probes and agents  
CC that bind to the antigen, especially monoclonal antibodies or  
CC equivalent polyclonal antibodies, are also used for diagnosis.

XX Sequence 368 AA;

alignment\_scores:

Quality: 713.00 Length: 143

Ratio: 5.057 Gaps: 0

Percent Similarity: 98.601 Percent Identity: 96.503

alignment\_block:

US-09-462-480-3 x AAW32384

Align seg 1/1 to: AAW32384 from: 1 to: 368

1 CTGCAGCAGGTGACGTCGTTGTTTTCAGCCAGGTGGCGCGCACCGCGCGG 50

|||||

226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGlyG1 242

51 CAACCCAGCCAGCAGGAGGAGCGGACATGGGCTGCTCGGCACCGATC 100

|||||

242 YAsnProAlaAspGluGluAlaGlnMetGlyLeuLeuGlyThrSerP 259

101 CGCTGTGCAACATCCGCTGCTGCTGATCAGCCAGCGCCGCGGCGG 150

|||||

259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275

151 GGCCTGCTGCGCGGAGTGCCTACCTGCGCAGGTGGTGGTGGTGGTGG 200

|||||

276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292

201 CACGCCGCTGATGCTCAGCTGATCGAAGAACCGGTTGCCCTCGGTGA 250

|||||

292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309

251 TGCCGGCGGCTGTTGCGGATCGTCGGTGACGGTGGCGCGCTCCGGTG 300

|||||

309 etProAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325

301 GGTCCGGGAGCAGTGGCGCAGGTTGCGCAATCCGCGCGCTCCACGAGCC 350

|||||

326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342

351 GGGTCTGTCGCGCGGCGCACGCTCGCGCAGGAGCGGTGAAGAAGACG 400

|||||

342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAspG 359

401 AGGACGACTGGGACGAGGACGACTGG 429

|||||

359 luAspAspTrpAspGluGluAspAspTrp 368

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW81705

seq\_documentation\_block:

ID AAW81705 standard; Protein; 368 AA.

XX



AC AAW81705;  
 XX 27-JAN-1999 (first entry)  
 XX  
 DE M. tuberculosis immunogenic polypeptide Tb37-FL.  
 XX  
 XX Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;  
 KW vaccine; pharmaceutical; infection; diagnosis.  
 XX  
 OS Mycobacterium tuberculosis.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 10  
 FT /label= unknown  
 XX  
 PN W09816646-A2.  
 XX  
 PD 23-APR-1998.  
 XX  
 PF 07-OCT-1997; 97WO-US18293.  
 XX  
 PR 13-MAR-1997; 97US-0818112.  
 PR 11-OCT-1996; 96US-0730510.  
 XX  
 XX (CORI-) CORIXA CORP.  
 PA  
 PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;  
 PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;  
 XX  
 DR WPI; 1998-261042/23.  
 XX  
 XX Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used  
 PT to develop products for the detection of M. tuberculosis infection  
 PT and for diagnosis, treatment and prevention of tuberculosis  
 XX  
 PS Example 3B; Page 137-138; 230pp; English.  
 XX  
 CC This sequence represents an immunogenic portion of a soluble  
 CC Mycobacterium tuberculosis (MT) antigen which can be used in a method  
 CC for inducing protective immunity against tuberculosis (TB). This  
 CC sequence can be formulated into vaccines and/or pharmaceutical  
 CC compositions for immunising against M. tuberculosis infection or may  
 CC be used for the diagnosis of tuberculosis.  
 XX  
 SQ Sequence 368 AA;  
 alignment\_scores:  
 Quality: 713.00 Length: 143  
 Ratio: 5.057 Gaps: 0  
 Percent Similarity: 98.601 Percent Identity: 96.503  
 alignment\_block:  
 US-09-462-480-3 x AAW81705 ..  
 Align seg 1/1 to: AAW81705 from: 1 to: 368  
 1 CTCGACGAGTGACGCTGTTGTCAGCCAGGTGGCGGCACCGCGCGG 50  
 |||||||||||||||||||||||||||||||||||||||||||||||  
 226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGly 242  
 51 CAACCCAGCCGACGAGGAGCCGCGCAGATGGCGCTGCTCGGCACCATC 100  
 |||||||||||||||||||||||||||||||||||||||||||||||  
 242 YasnProAlaAspGluAlaGlnMetGlyLeuLeuGlyThrSerP 259  
 101 CGCTGTGCAACCATCCGCTGGCTGGTGATCAGGCCCGCCCGCGCGG 150  
 |||||||||||||||||||||||||||||||||||||||||||||||  
 259 roLeuSerAsnHisProLeuAlaGlySerGlyProSerAlaGlyAla 275  
 151 GGCTGCTGCGCGGAGTGCTACCTGCGCAGGTGGGTGCTGACCCG 200  
 |||||||||||||||||||||||||||||||||||||||||||||||  
 276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292

201 CACGCGGCTGATCTCAGCTGATCGAAAAAGCCGTTGCCCTCGGTGA 250  
 |||||||||||||||||||||||||||||||||||||||||||||||  
 292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309  
 251 TGCCGCGGCTGTTGCGGATCOTCGGTGACGGTGGCGCGCTCGCGTG 300  
 |||||||||||||||||||||||||||||||||||||||||||||||  
 309 etProAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325  
 301 GGTCCGCGGAGCGATGGCCAGGTTCGCAATCGCGGCTCCACGACCC 350  
 ||| ||||||||||||||||||||||||||||||||||||||||  
 326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342  
 351 GGCTCTGGTCGCGCGCCGCTCGCGCAGGAGCGGTGAAGAAGACGACG 400  
 |||||||||||||||||||||||||||||||||||||||||||||||  
 342 OGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAspG 359  
 401 AGGACGACTGGGACGAGGAGGAGGACTGG 429  
 ||||||||||||||||||||||||||||||||||||||||  
 359 luAspAspTrpAspGluGluAspAspTrp 368  
 seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW64338  
 seq\_documentation\_block:  
 ID AAW64338 standard; Protein; 368 AA.  
 XX  
 AC AAW64338;  
 XX  
 DT 09-NOV-1998 (first entry)  
 XX  
 DE Mycobacterium tuberculosis antigen Tb37-FL.  
 KW Tuberculosis; infection; diagnosis; antigen; Tb37-FL.  
 XX  
 OS Mycobacterium tuberculosis strain H37Rv.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 10  
 FT /note= "unidentified"  
 XX  
 PN W09816645-A2.  
 XX  
 PD 23-APR-1998.  
 XX  
 PF 07-OCT-1997; 97WO-US18214.  
 XX  
 PR 13-MAR-1997; 97US-0818111.  
 PR 11-OCT-1996; 96US-0729622.  
 XX  
 XX (CORI-) CORIXA CORP.  
 PA  
 PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;  
 PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;  
 XX  
 DR WPI; 1998-251292/22.  
 XX  
 PT New isolated Mycobacterium tuberculosis polypeptides and DNA - used  
 PT to develop products for the detection of M. tuberculosis infection  
 XX and diagnosis of tuberculosis  
 PS Example 3; Page 143-144; 250pp; English.  
 XX  
 CC This polypeptide comprises Mycobacterium tuberculosis antigen  
 CC Tb37-FL. It is encoded by genomic DNA isolated from a M.  
 CC tuberculosis strain H37Rv genomic library using a probe from  
 CC clone Tb38-1 (see AAW44384). The invention relates to compositions  
 CC and methods for diagnosing tuberculosis. It provides polypeptides  
 CC (see AAW64291-W64379) comprising an antigenic portion of a soluble  
 CC M. tuberculosis antigen, or an immunogenic portion of an M.  
 CC tuberculosis antigen, as well as DNA sequences encoding such  
 CC polypeptides, recombinant expression vectors and transformed or  
 CC transfected host cells. Also claimed are methods and diagnostic  
 CC kits for detecting M. tuberculosis infection in a patient using  
 CC these polypeptides, antibodies or oligonucleotide probes and

CC primers, for the diagnosis of tuberculosis.

XX Sequence 368 AA;

alignment\_scores:

Quality: 713.00 Length: 143  
Ratio: 5.057 Gaps: 0  
Percent Similarity: 98.601 Percent Identity: 96.503

alignment\_block:

US-09-462-480-3 x AAW64338 ..

Align seg 1/1 to: AAW64338 from: 1 to: 368

```
1 CTGCAGCAGGTGACGTCGTTGTTTCAGCCAGGTGGCGGCACCGCGCGG 50
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGlyG 242
51 CAACCCAGCCGACGAGGAAGCCGCGAGATGGGCTGCTCGGCACCAATC 100
242 YAsnProAlaAspGluAlaGlnMetGlyLeuLeuGlyThrSerP 259
101 CGCTGTGCAACCATCCGCTGCTGGTGGATCAGCCGCCAGCGCGGCGG 150
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275
151 GGCCTGCTGCGCGCGGAGTGCCTACCTGGCGCAGGTGGTGGTGGTGG 200
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292
201 CACGCCGCTGATGCTCAGCTGATCGAAGCCGGTGGCCCTCGGTGA 250
292 gThrProLeuMetSerGlnLeuLeuGluLysProValAlaProSerValM 309
251 TCCCGCGGCTGTTGCGGATCGTCGATGACGGGTGGCGGCTCGGCTG 300
309 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
301 GGTCTGCTGCGCGGAGGTGGCGGCTGCAATCGCGGCTCCACCGCC 350
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrAr 342
351 GGTCTGCTGCGCGGAGGTGGCGGCTGCAATCGCGGCTCCACCGCC 400
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluAspAspG 359
401 AGGACGACTGGGACGAGGACGACTGG 429
359 luAspAspTrpAspGluGluAspAspTrp 368
```

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.AAY39135

seq\_documentation\_block:

ID AAY39135 standard; Protein; 368 AA.

XX AC AAY39135;

XX DT 05-NOV-1999 (first entry)

XX DE M. tuberculosis antigen Tb37-FL amino acid sequence.

XX KW Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;  
KW immunotherapy; diagnosis; immunisation; vaccine; infection;  
KW immune response; skin test.

XX OS Mycobacterium tuberculosis.

XX PN W09942076-A2.

XX PD 26-AUG-1999.

XX PF 17-FEB-1999; 99WO-US03268.

XX

PR 05-MAY-1998; 98US-0072967.  
PR 18-FEB-1998; 98US-0025197.

XX (CORI-) CORIXA CORP.

XX Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;  
PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedwick TS;  
XX WPI; 1999-527409/44.

XX New antigens from Mycobacterium tuberculosis useful in diagnostic  
PT skin tests and protective or therapeutic vaccines or compositions  
XX Example 3; Page 132-133; 299pp; English.

XX The present invention describes polypeptides comprising an immunogenic  
CC part of a Mycobacterium tuberculosis antigen (Ag). Also described  
CC are vaccines and fusion protein containing M. tuberculosis Ag's.  
CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and  
CC other polypeptides fragments, can be used in pharmaceutical compositions  
CC or vaccines to generate a protective or therapeutic immune response to  
CC M. tuberculosis and as reagents in skin tests for diagnosis of  
CC tuberculosis. Ag can induce proliferation of, or cytokine secretion  
CC by, T, B or natural killer cells and/or macrophages in  
CC tuberculosis-immune subjects. AAZ19249 to AAZ19460 and AAY39083 to  
CC AAY39225 are used in the exemplification of the present invention.

XX Sequence 368 AA;

alignment\_scores:

Quality: 713.00 Length: 143  
Ratio: 5.057 Gaps: 0  
Percent Similarity: 98.601 Percent Identity: 96.503

alignment\_block:

US-09-462-480-3 x AAY39135 ..

Align seg 1/1 to: AAY39135 from: 1 to: 368

```
1 CTGCAGCAGGTGACGTCGTTGTTTCAGCCAGGTGGCGGCACCGCGCGG 50
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGlyG 242
51 CAACCCAGCCGACGAGGAAGCCGCGAGATGGGCTGCTCGGCACCAATC 100
242 YAsnProAlaAspGluAlaGlnMetGlyLeuLeuGlyThrSerP 259
101 CGCTGTGCAACCATCCGCTGCTGGTGGATCAGCCGCCAGCGCGGCGG 150
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275
151 GGCCTGCTGCGCGGAGTGCCTACCTGGCGCAGGTGGTGGTGGTGGTGG 200
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292
201 CACGCCGCTGATGCTCAGCTGATCGAAGCCGGTGGCCCTCGGTGA 250
292 gThrProLeuMetSerGlnLeuLeuGluLysProValAlaProSerValM 309
251 TCCCGCGGCTGTTGCGGATCGTCGATGACGGGTGGCGGCTCCCGGTG 300
309 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
301 GGTCTGCTGCGCGGAGGTGGCGGCTGCAATCGCGGCTCCACCGCC 350
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrAr 342
351 GGTCTGCTGCGCGGAGGTGGCGGCTGCAATCGCGGCTCCACCGCC 400
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluAspAspG 359
```

401 AGGACGACTGGGACGAGAGGACGACTGG 429  
 359 luAspAspTirpAspGluGluAspAspTrp 368

seq\_name: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AA138992

seq\_documentation\_block:

ID AAY38992 standard; Protein: 368 AA.

AC AAY38992;

DT 05-NOV-1999 (first entry)

DE M. tuberculosis recombinant antigen protein Tb37-FL.

KW Antigen; diagnosis; detection; infection; antibody; immunisation;

KW vaccine; immunity.

OS Mycobacterium tuberculosis.

XX WO9942118-A2.

PD 26-AUG-1999.

PF 17-FEB-1999; 99WO-US03265.

PR 05-MAY-1998; 98US-0072596.

PR 18-FEB-1998; 98US-0024753.

PA (CORI-) CORIXA CORP.

XX

PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;

PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedwick TS;

XX WPI; 1999-527416/44.

PT New polypeptide comprising antigenic portions of M. tuberculosis

XX Example 3; Page 177-179; 323pp; English.

CC This invention describes novel recombinant antigens and their encoding  
 CC nucleic acids derived from Mycobacterium tuberculosis. The novel  
 CC polypeptides are useful for detecting M. tuberculosis infection in a  
 CC biological sample by detecting antibodies which bind with the  
 CC polypeptides, and are useful as vaccines for immunizing against  
 CC M. tuberculosis infection. The new detection methods are needed as  
 CC current vaccination strategies do not provide 100% immunity.

XX Sequence 368 AA;

alignment\_scores:

Quality: 713.00 Length: 143  
 Ratio: 5.057 Gaps: 0  
 Percent Similarity: 98.601 Percent Identity: 96.503

alignment\_block:

US-09-462-480-3 x AAY38992 ..

Align seg 1/1 to: AAY38992 from: 1 to: 368

1 CTGCAGCAGGTGACGTCGTTGTCAGCCAGGTGGCGCGCACCGCGCGG 50

226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGlyG1 242

51 CAACCCAGCCGACGAGGAGCGCGCAGATGGGCTGCTCGCGCACGATC 100

242 YAsnProAlaAspGluGluAlaAlaGlnMetGlyLeuLeuGlyThrSerP 259

101 CGCTGTCAACATCCGTCGTGGTGGATCAGGCCCGCGCGCGCGGCGG 150

259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275

151 GGCCTGCTCGCGCGAGTCGCTACCTGCCGAGGTGGTGGTTCGTGACCCG 200  
 276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292  
 201 CACGCCGTGATGTCACGTGATCGAAAGCCGGTTCGCCCTCGGTGA 250  
 292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309  
 251 TCCGCGCGCTGTTCCGCGATCGTGGTGGCGGTGGCGCGCTCCGCTG 300  
 309 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325  
 301 GGTCCGGAGGATGGCCAGGTTTCGCAATCCGCGGCTCCACACGCC 350  
 356 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342  
 351 GGTCTGTGTCGCGCGCGCACCGCTCGCGCAGGAGCGTGAAGAAGACGAG 400  
 342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAspG 359  
 401 AGGACGACTGGGACGAGGACGACTGG 429  
 359 luAspAspTirpAspGluGluAspAspTrp 368

seq\_name: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT: AAB19844

seq\_documentation\_block:

ID AAB19844 standard; Protein: 368 AA.

XX AAB19844;

XX 05-MAR-2001 (first entry)

DE Mycobacterium tuberculosis protein MTBN3.

XX MTBN3; tuberculosis; BCG; vaccine; infection; diagnosis.

XX Mycobacterium tuberculosis.

XX WO2000066157-A1.

XX 09-NOV-2000.

XX 04-MAY-2000; 2000WO-US12257.

XX 04-MAY-1999; 99US-0132505.

XX (PUBL-) PUBLIC HEALTH RES INST NEW YORK.

XX Gennaro ML;

XX WPI; 2001-007153/01.

XX N-PSDB; AAA89037.

Novel polypeptide encoded by open reading frames present in  
 Mycobacterium tuberculosis genome and not by the BCG strain of M.  
 bovis, useful as vaccine and for diagnosing tuberculosis infection

Claim 11; Fig 1; 35pp; English.

XX The present sequence is that of the Mycobacterium tuberculosis  
 CC MTBN3 protein. This is 1 of 8 proteins, i.e. MTBN1-8 (see  
 CC AAB19842-49), encoded by 8 open reading frames (see AAA89035-42)  
 CC identified as being present in the genome of M. tuberculosis but  
 CC absent from the genome of the BCG strain of Mycobacterium bovis.  
 CC MTBN1-8 represent reagents that are useful in discriminating between  
 CC M. tuberculosis and BCG and, in particular, for diagnostic methods  
 CC which discriminate between exposure of a subject to M. tuberculosis  
 CC and vaccination with BCG. The invention features these MTBN  
 CC polypeptides, functional fragments of them, DNA encoding them,  
 CC vectors, transformed cells, and diagnostic, therapeutic, and  
 CC prophylactic (vaccine) methods, including genetic vaccination  
 CC methods.

XX  
SQ Sequence 368 AA;

## alignment\_scores:

Quality: 713.00 Length: 143  
Ratio: 5.057 Gaps: 0  
Percent Similarity: 98.601 Percent Identity: 96.503

## alignment\_block:

US-09-462-480-3 x AAB19844

Align seg 1/1 to: AAB19844 from: 1 to: 368

```

1 CTGCAGCAGGTGACGTCGCTGTTTCAGCAGGTGGCGGCACCGGGCGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGlyG1 242
51 CAACCCAGCCGACGAGGAGGAGCGCGCAGATGGGCTGTCGGCACCATC 100
|||||
242 yAsnProAlaAspGluGluAlaAlaGlnMetGlyLeuLeuGlyThrSerP 259
101 CGCTGCTCAACCATCCGCTGGCTGGTGCATCAGCCGCCAGCGCGGCGG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275
151 GGCCTGCTCGCGCGGAGTCGCTACTCGCGCAGGTGGTGGTTCGACCCG 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292
201 CAGCGCGCTGATGTCATGCTAGTCATCGAAAGCCGGTTCGCCCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309
251 TGCCGGCGGCTGTTCCCGGATCGTCGGTGACGGTGCGCGCTCCGGTG 300
|||||
309 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
301 GGTCCGGGAGCGATGGCGCAGGTTTCGCAATCCGCGCGCTCCACGACCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
351 GGGTCTGCTCGCGCGGCGCACGCTCGCGCAGGAGCGGTGAAGAAGACGAC 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAspG 359
401 AGGACGACTGGGACGAGGAGGACGACTGG 429
|||||
359 luAspAspTrpAspGluGluAspAspTrp 368

```

seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:AAW72929

## seq\_documentation\_block:

ID\_AAW72929 standard; Protein; 371 AA.

XX

AC AAW72929;

XX

DT 21-JAN-1999 (first entry)

XX

DE Mycobacterium tuberculosis antigen RD1-ORE5.

XX

KW Mycobacterium tuberculosis; antigen; vaccine; immunological;

XX

OS Mycobacterium tuberculosis.

XX

PN WO9844119-A1.

XX

PD 08-OCT-1998.

XX

PF 01-APR-1998; 98WO-DK00132.

XX

XX 05-JAN-1998; 98US-0070488.

PR 02-APR-1997; 97DK-0000376.  
PR 18-APR-1997; 97US-0044624.  
PR 10-NOV-1997; 97DK-0001277.

XX  
XX (STAT-) STATENS SERUM INST.

XX Andersen P, Florio W, Nielsen R, Oettinger T, Rasmussen PB;

PI Rosenkrands I, Weidling K;

XX  
XX WPI: 1998-542705/46.

DR N-PSDB; AAV63939.

XX

PT New isolated mycobacteria polypeptides and nucleic acids - used for  
PT developing products for the diagnosis of or vaccination against  
PT mycobacterial infections, particularly tuberculosis

XX Claim 1; Page 200-202; 163pp; English.

CC The present sequence represents a Mycobacterium tuberculosis protein.  
CC Products from the present invention, which describes protein fragments  
CC and nucleic acid fragments derived from M.tuberculosis, can be used in  
CC the detection of and prevention of mycobacterial infections. In  
CC particular, the proteins and nucleic acids can be used for the diagnosis  
CC of or vaccination against tuberculosis caused by M. tuberculosis,  
CC M. africanum or M. bovis.

XX  
SQ Sequence 371 AA;

## alignment\_scores:

Quality: 713.00 Length: 143  
Ratio: 5.057 Gaps: 0  
Percent Similarity: 98.601 Percent Identity: 96.503

## alignment\_block:

US-09-462-480-3 x AAW72929

Align seg 1/1 to: AAW72929 from: 1 to: 371

```

1 CTGCAGCAGGTGACGTCGCTGTTTCAGCAGGTGGCGGCACCGGGCGG 50
|||||
229 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGlyG1 245
51 CAACCCAGCCGACGAGGAGGAGCGCGCAGATGGGCTGTCGGCACCATC 100
|||||
245 yAsnProAlaAspGluGluAlaAlaGlnMetGlyLeuLeuGlyThrSerP 262
101 CGCTGCTCAACCATCCGCTGGTGGTGCATCAGCCGCCAGCGCGGCGG 150
|||||
262 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 278
151 GGCCTGCTCGCGCGGAGTCGCTACTCGCGCAGGTGGTGGTTCGACCCG 200
|||||
279 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 295
201 CAGCGCGCTGATGTCATGCTAGTCATCGAAAGCCGGTTCGCCCTCGGTGA 250
|||||
295 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 312
251 TGCCGGCGGCTGTTCCCGGATCGTCGGTGACGGTGCGCGCTCCGGTG 300
|||||
312 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 328
301 GGTCCGGGAGCGATGGCGCAGGTTTCGCAATCCGCGCGCTCCACGACCC 350
|||||
329 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 345
351 GGGTCTGCTCGCGCGGCGCACGCTCGCGCAGGAGCGGTGAAGAAGACGAC 400
|||||
345 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAspG 362
401 AGGACGACTGGGACGAGGAGGACGACTGG 429
|||||

```

362 luaspaspTrpaspGluGluAspAspTrp 371

seq\_name: /SIDSl/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AA1999

seq\_documentation\_block:

ID AAY21946 standard; Protein; 371 AA.

AC AAY21946;

DT 06-SEP-1999 (first entry)

DE Amino acid sequence of antigen RD1-ORF5.

XX Immunogenic; Mycobacterium tuberculosis; Immune response; Infection;  
KW tuberculosis; fusion polypeptide; T-cell epitope; ESAT-6; MPT59; TB;  
KW pharmaceutical; vaccination; M. africanum; M. bovis; CFP7A; CFP30A;  
KW CFP7B; CFP19; CFP27; CFP30A; RD1-ORF; CFP10A; CFP16; CFP19; CFP23;  
KW CFP25A; CFP30B; CFP7B.

XX Mycobacterium tuberculosis.

OS WO9924577-A1.

PN 20-MAY-1999.

PD 08-OCT-1998; 98WO-DK00438.

PR 01-APR-1998; 98WO-DK00132.

PR 10-NOV-1997; 97DK-0001277.

PR 05-JAN-1998; 98US-0070488.

XX (STAT-) STATENS SERUM INST.

PA Andersen P, Skjot R;

XX WPI; 1999-347282/29.

DR N-PSDB; AAX81046.

XX New immunogenic fragment of Mycobacterium tuberculosis.

PT Example 2; Page 219-220; 265pp; English.

XX The invention describes a substantially pure immunogenic polypeptide  
CC fragment (I) from Mycobacterium tuberculosis that is able to evoke a  
CC protective immune response against infections by mycobacteria belonging  
CC to the tuberculosis complex. The invention provides a (1) fusion  
CC polypeptide comprising at least one polypeptide fragment (1) and at least  
CC one fusion partner; (2) a fusion polypeptide fragment comprising a T-cell  
CC epitope from M. tuberculosis protein ESAT-6, or MPT59 and a second  
CC different amino acid sequence from M. tuberculosis, and/or including a  
CC sequence which protects the first amino acid sequence from in vivo  
CC degradation or post-translational processing; (3) a nucleic acid fragment  
CC that encodes the above polypeptides. The polypeptides and nucleic acid  
CC are useful as pharmaceuticals, for diagnosis of and as antigens for  
CC vaccination against TB caused by Mycobacterium tuberculosis, africanum or  
CC bovis. The polypeptides are also useful for diagnosing ongoing or  
CC previous sensitization in an animal with bacteria belonging to the  
CC tuberculosis complex. The invention also describes the use of CFP7A or  
CC CFP30A or a T-cell epitope of for the induction of a strong immune  
CC response in a mammal; use of CFP7B, CFP19 or MPT59-ESAT6 or a T-cell  
CC epitope of for diagnosis of TB in a mammal by performing a DTH type skin  
CC test; use of CFP27, CFP30A, RD1-ORF2, RD1-ORF3, RD1-ORF5, MPT59-ESAT6,  
CC ESAT6-MPT59, CFP10A, CFP16, CFP19, CFP23, CFP25A, CFP30B, CFP7B or a T-  
CC cell epitope of for the preparation of an immunological composition; and  
CC for the preparation of a subunit vaccine.

XX SQ Sequence 371 AA;

alignment\_scores:

Quality:	713.00	Length:	143
Ratio:	5.057	Gaps:	0
Percent Similarity:	98.601	Percent Identity:	96.503

alignment\_block:

US-09-462-480-3 x AAY21946

Align seg 1/1 to: AAY21946 from: 1 to: 371

1 CTGCAGCAGGTGACGTCGTTGTTTCAGCCAGGTGGCGGCGCACCGCGGCGG 50  
|||||  
229 LeuGlnGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGly 245  
51 CAACCCAGCCGACGAGGAGCCGCGATGGCGCTGCTCGGCACAGTC 100  
245 yasnProAlaAspGluGluAlaAlaGlnMetGlyLeuLeuGlyThrSerP 262  
101 CGCTGTCGAACCATCCGCTGGCTGGTGGATCAGGCCCGCCAGCGCGCGG 150  
|||||  
262 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 278  
151 GGCCTGCTGCGCGGAGTGCCTACTCTGGCGCAGGTGGGTGCTTGACCCG 200  
279 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 295  
201 CAGCGCGCTGATGCTCAGCTGATCGAAAGCCGGTGGCCCTCGGTGA 250  
295 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValI 312  
251 TCGCGCGGCTGTCGCGGATCGTCGGTGACGGTGGCGCGCTCGCGTG 300  
312 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 328  
301 GGTCCGGAGCGATGGCGCAGGTTTCGCAATCCGGCGCTCCACCGCCC 350  
329 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 345  
351 GGGTCTGGTTCGCGCGGCGCACCCCTCGCGCAGGAGCGTGAAGAGAGCAGC 400  
345 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAspG 362  
401 AGGACGACTGGGACGAAGAGGACGACTGG 429  
362 luaspaspTrpaspGluGluAspAspTrp 371

seq\_name: /SIDSl/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW72928

seq\_documentation\_block:

ID AAW72928 standard; Protein; 139 AA.

XX AC AAW72928;

XX 21-JAN-1999 (first entry)

XX Mycobacterium tuberculosis antigen RD1-ORF4.

XX Mycobacterium tuberculosis; antigen; vaccine; immunological;  
KW immunogen; infection.

XX Mycobacterium tuberculosis.

PN WO9844119-A1.

PD 08-OCT-1998.

XX 01-APR-1998; 98WO-DK00132.

XX 05-JAN-1998; 98US-0070488.

PR 02-APR-1997; 97DK-0000376.

PR 18-APR-1997; 97US-0044624.

PR 10-NOV-1997; 97DK-0001277.

XX (STAT-) STATENS SERUM INST.

XX Andersen P, Florio W, Nielsen R, Oettinger T, Rasmussen PB;

PI Rosenkrands I, Weidigh K;

XX WPI; 1998-542705/46.  
DR N-PSDB; AAV63938.  
XX New isolated mycobacteria polypeptides and nucleic acids - used for  
PT developing products for the diagnosis of or vaccination against  
PT mycobacterial infections, particularly tuberculosis  
XX Claim 1; Page 198; 163pp; English.  
XX The present sequence represents a Mycobacterium tuberculosis protein.  
CC Products from the present invention, which describes protein fragments  
CC and nucleic acid fragments derived from M. tuberculosis, can be used in  
CC the detection of and prevention of mycobacterial infections. In  
CC particular, the proteins and nucleic acids can be used for the diagnosis  
CC of or vaccination against tuberculosis caused by M. tuberculosis,  
CC M. africanum or M. bovis.  
XX  
SQ Sequence 139 AA;

alignment\_scores:  
Quality: 373.00 Length: 67  
Ratio: 5.567 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 98.507

alignment\_block:  
US-09-462-480-3/rev x AAW72928 ..  
Align seg 1/1 to: AAW72928 from: 1 to: 139

203 GTCCGGGTCAACAGCCACTGCGCAGGTAGCGACTCCGCGGCAGCAG 154  
1 MetArgValAsnAspProAlaProGlySerAspAlaArgSerAr 17  
153 GCCCGCGCGCGCTGGCGCTGATCCACAGCCAGCGGATGTTGCACA 104  
17 gProAlaProAlaLeuGlyProAspProAlaSerGlyTrpPheAsps 34  
103 GCGGACTGTGTCGAGCAGCCCATCTGCGCGCTTCCCTCGCTGGG 54  
34 erGlyLeuValProSerArgProIleCysAlaAlaSerSerAlaGly 50  
53 TTCCCGCGCGCTGGCGCGCCACTGGCTGACACAGCGTACCTGCTG 4  
51 LeuProProValProProThrTrpLeuAsnAspValThrCysCys 67  
3 C 3  
67 s 67

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAV21945

seq\_documentation\_block:  
ID AAV21945 standard; Protein; 139 AA.

XX AAV21945;  
AC  
XX  
DT 06-SEP-1999 (first entry)  
XX  
DE Amino acid sequence of antigen RDI-ORF4.  
XX Immunogenic; Mycobacterium tuberculosis; immune response; infection;  
KW tuberculosis; fusion polypeptide; T-cell epitope; ESAT-6; MPT59; TB;  
KW pharmaceutical; vaccination; M. africanum; M. bovis; CFP7A; CFP30A;  
KW CFP7B; CFP19; CFP27; CFP30A; RDI-ORF; CFP10A; CFP16; CFP19; CFP23;  
KW CFP25A; CFP30B; CFP7B.  
XX Mycobacterium tuberculosis.  
OS  
XX W09924577-A1.  
PN  
XX  
PD 20-MAY-1999.

XX 08-OCT-1998; 98WO-DK00438.  
PF  
XX  
PR 01-APR-1998; 98WO-DK00132.  
PR 10-NOV-1997; 97DK-0001277.  
PR 05-JAN-1998; 98US-0070488.  
XX  
PA (STAT-) STATENS SERUM INST.  
XX  
PI Andersen P, Skjot R;  
XX  
DR WPI; 1999-347282/29.  
DR N-PSDB; AAX81045.  
XX  
PT New immunogenic fragment of Mycobacterium tuberculosis  
XX  
PS Example 2; Page 216; 265pp; English.  
XX  
CC The invention describes a substantially pure immunogenic polypeptide  
CC fragment (I) from Mycobacterium tuberculosis that is able to evoke a  
CC protective immune response against infections by mycobacteria belonging  
CC to the tuberculosis complex. The invention provides a (1) fusion  
CC polypeptide comprising at least one polypeptide fragment (I) and at least  
CC one fusion partner; (2) a fusion polypeptide fragment comprising a T-cell  
CC epitope from M. tuberculosis protein ESAT-6, or MPT59 and a second  
CC different amino acid sequence from M. tuberculosis, and/or including a  
CC sequence which protects the first amino acid sequence from in vivo  
CC degradation or post-translational processing; (3) a nucleic acid fragment  
CC that encodes the above polypeptides. The polypeptides and nucleic acid  
CC are useful as pharmaceuticals, for diagnosis of and as antigens for  
CC vaccination against TB caused by Mycobacterium tuberculosis, africanum or  
CC bovis. The polypeptides are also useful for diagnosing ongoing or  
CC previous sensitization in an animal with bacteria belonging to the  
CC tuberculosis complex. The invention also describes the use of CFP7A or  
CC CFP30A or a T-cell epitope of for the induction of a strong immune  
CC response in a mammal; use of CFP7B, CFP19 or MPT59-ESAT6 or a T-cell  
CC epitope for diagnosis of TB in a mammal by performing a DTH type skin  
CC test; use of CFP27, CFP30A, RDI-ORF2, RDI-ORF3, RDI-ORF5, ESAT6,  
CC ESAT6-MPT59, CFP10A, CFP16, CFP19, CFP23, CFP25A, CFP30B, CFP7B or a T-  
CC cell epitope of for the preparation of an immunological composition; and  
CC for the preparation of a subunit vaccine.  
XX  
SQ Sequence 139 AA;

alignment\_scores:  
Quality: 373.00 Length: 67  
Ratio: 5.567 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 98.507

alignment\_block:

US-09-462-480-3/rev x AAY21945 ..

Align seg 1/1 to: AAY21945 from: 1 to: 139

203 GTCCGGGTCAACAGCCACTGCGCAGGTAGCGACTCCGCGGCAGCAG 154  
1 MetArgValAsnAspProAlaProGlySerAspAlaArgSerAr 17  
153 GCCCGCGCGCGCTGGCGCTGATCCACAGCCAGCGGATGTTGCACA 104  
17 gProAlaProAlaLeuGlyProAspProAlaSerGlyTrpPheAsps 34  
103 GCGGACTGTGTCGAGCAGCCCATCTGCGCGCTTCCCTCGCTGGG 54  
34 erGlyLeuValProSerArgProIleCysAlaAlaSerSerAlaGly 50  
53 TTCCCGCGCGCTGGCGCGCCACTGGCTGACACAGCGTACCTGCTG 4  
51 LeuProProValProProThrTrpLeuAsnAspValThrCysCys 67  
3 C 3



ARB54248;  
09-MAR-2001 (first entry)  
Human pancreatic cancer antigen protein sequence SEQ ID NO:700.  
Human; pancreas; pancreatic cancer; pancreatic cancer antigen;  
detection; diagnosis; identification; cytostatic; neuroprotective;  
nootropic; immunomodulatory; relaxant; contraceptive; gynaecological;  
antiinflammatory; cardiant; gene therapy; chromosome mapping;  
linkage analysis; tissue identification; tissue typing; forensic;  
neural; immune system; muscular; reproductive; gastrointestinal;  
pulmonary; cardiovascular; renal; proliferative.  
Homo sapiens.  
WO200055320-A1.  
21-SEP-2000.  
08-MAR-2000; 2000WO-US05989.  
12-MAR-1999; 99US-0124270.  
(HUMA-) HUMAN GENOME SCI INC.  
Rosen CA, Ruben SM;  
WPI; 2000-579444/54.  
N-PSDB; AAC99013.  
New nucleic acid that is a pancreatic cancer antigen for preventing,  
treating, or ameliorating a medical condition, particular pancreatic  
cancer, or for use in assays for diagnosing a pathological condition -  
Claim 11; Page 1137; 1379pp; English.  
AAC98773 to AAC99231 encode the human pancreatic cancer associated  
proteins, called pancreatic cancer antigens, given in AAB54008 to  
AAB54466. The human pancreatic cancer antigens have cytostatic,  
neuroprotective, nootropic, immunomodulatory, relaxant, contraceptive,  
gynaecological, cardiant and antiinflammatory activities, and can be used  
in gene therapy. The polynucleotide and proteins can be used for  
preventing, treating, or ameliorating a medical condition or in assays  
for diagnosing a pathological condition or a susceptibility to one in a  
subject. Binding partners to the proteins and the activity of the  
proteins can be identified. The pancreatic cancer antigens can be used to  
detect, treat or prevent pancreatic disorders, especially cancer.  
Agonists and antagonists to the antigens can be screened for. The  
pancreatic cancer antigen polynucleotides can be used to design nucleic  
acid hybridisation probes that can be used in chromosome mapping, linkage  
analysis, tissue identification and/or typing and a variety of forensic  
and diagnostic methods. The proteins can be used to generate antibodies  
which are used to purify, detect and target the polypeptides, including  
both in vivo and in vitro diagnostic and therapeutic methods. The  
proteins can be used to treat or prevent neural, immune system, muscular,  
reproductive, gastrointestinal, pulmonary, cardiovascular, renal or  
proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent  
sequences used in the exemplification of the present invention.  
Sequence 200 AA;  
alignment\_scores:  
Quality: 145-50 Length: 163  
Ratio: 2.021 Gaps: 10  
Percent Similarity: 44.172 Percent Identity: 33.742  
alignment\_block:  
US-09-462-480-3/rev x AAB54248 ..  
Align seg 1/1 to: AAB54248 from: 1 to: 200

472 CGGCCCGCGGTGGCCGGGAAGTCTGTGTGATTACGGGAGCTCACCAGTGC 423  
||||| |||  
12 ArgProGlyProPro..... 16  
422 TCCTCTCTCGTCCACGTCGT.....CCTC 400  
||||| |||||  
17 .ProLeuProProSerSerSerGlyAsnAspGluThrProArgLeuProG 33  
399 GTCTGTCTTCTTTCACCTCTCTGCGCGAGGGTCCGCGCGACGACGACCG 350  
||||| ||| : : : : : |||||  
33 InArgAsnLeu...SerLeuSerSerThrProProLeuProSerPro 48  
349 GGCTGGTGGAGCCGCGGATTGCGAACCTGGCCCATCGCTCCCGGACCC 300  
||| : : : ||||| |||  
49 Gly.....ArgSerGlyPro...LeuProPro 57  
299 ACCGGAGCGGCCACCGTCACCGACGATCGGGCAACAGCGCGCGGAT 250  
||||| : : : ||| ||| : : : |||  
57 oProSerGluArg...ProProProProValArgAspProProGlyArg 73  
249 CA.....CCGAGGGGGCAACCGGCTTTTCGATCAGCTGAGACATCA 209  
|| ||| ||||| : : : : : |||||  
73 eGlyProLeuProProProProProValSerArgAsnGlySerThrSer 89  
208 GCGCGCTGCGGTCAACGACCCACCTGGCGCA.....GGTAGCGAC 168  
||| : : : ||| ||| |||  
90 ArgAlaLeuProAlaThr...ProGlnLeuProSerArgSerGlyValAsp 105  
167 TCCGCGCGCAGCGCGCGCGCGCGCTGGGGCTGATCCACGACGACG 118  
106 SerProArgSerGlyProArgProProLeuProProAspArgPro..... 120  
117 CGGATGTTGACAGCGGACTGGTCCGAGCAGGCCCATCTCGCGCGGCTT 68  
120 ..... 120  
67 CTCTGCTGCGGTGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 30  
||||| ||||| ||||| |||||  
121 .....SerAlaGlyAlaProProProProProProSer 131  
seq\_name: /SDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:AAM79841  
seq\_documentation\_block:  
ID AAM79841 standard; Protein; 1092 AA.  
XX  
AC AAM79841;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human protein SEQ ID NO 3487.  
XX  
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorder; arthritis; inflammation.  
XX  
OS Homo sapiens.  
XX  
PN WO200157190-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 05-FEB-2001; 2001WO-US04098.  
XX  
PR 03-FEB-2000; 2000US-0496914.  
PR 27-APR-2000; 2000US-0560875.  
PR 20-JUN-2000; 2000US-0598075.  
PR 19-JUL-2000; 2000US-0620325.  
PR 01-SEP-2000; 2000US-0654936.  
PR 15-SEP-2000; 2000US-0663561.  
PR 20-OCT-2000; 2000US-0693325.  
PR 30-NOV-2000; 2000US-0728422.  
XX



OPTIONAL FORM NO. 10

Category	Item	Value	Unit
Agriculture	Wheat	1200	kg
	Rice	800	kg
	Corn	500	kg
	Soybeans	300	kg
	Cotton	200	kg
	Vegetables	100	kg
	Fruits	50	kg
	Flowers	20	kg
	Herbs	10	kg
	Medicines	5	kg
Livestock	Cattle	100	kg
	Pigs	80	kg
	Sheep	60	kg
	Goats	40	kg
	Chickens	20	kg
	Ducks	10	kg
	Geese	5	kg
	Bees	2	kg
	Ants	1	kg
	Termites	0.5	kg
Fishing	Shrimp	100	kg
	Crab	80	kg
	Clam	60	kg
	Oyster	40	kg
	Scallop	20	kg
	Sea urchin	10	kg
	Starfish	5	kg
	Sea cucumber	2	kg
	Sea slug	1	kg
	Sea anemone	0.5	kg
Hunting	Deer	100	kg
	Wild boar	80	kg
	Wild cat	60	kg
	Wild dog	40	kg
	Wild rabbit	20	kg
	Wild bird	10	kg
	Wild fish	5	kg
	Wild insect	2	kg
	Wild plant	1	kg
	Wild fungus	0.5	kg

32

alignment_scores:	
Quality:	140.00
Ratio:	2.000
Percent Similarity:	42.945
Percent Identity:	32.515
Gaps:	9
Length:	163

```
alignment_block:
US-09-462-480-3/rev x AAY24091
Align seg 1/1 to: AAY24091 from: 1 to: 504

472 CGGCGCGGTGGCGGGAAGTCTGTGTCATTACGGGAGCTCACCAGTGG 423
|||||
315 ArgProGlyProPro..... 319
422 TCCTCTCGTCCAGTCGT.....CCTC 400
|||||
320 .ProLeuProProSerSerGlyAsnAspGluThrProArgLeuProG 336
|||||
399 GTCGCTCTTTCACGCTCTGCGGAGGGTCCGCGCGACGACGACCG 350
|||||
336 InArgAsnLeu...SerLeuSerSerThrProProLeuProSerPro 351
|||||
349 GGCTGGTGAGCGCGGATTGCGAACCTGGCCCATCGTCCCGGACCC 300
|||||
352 Gly.....ArgSerGlyPro...LeuProProPr 360
299 ACCGGAGCGCGCCACCGTCTACCGACGATCCGCAACAGCGCGGCAT 250
|||||
360 oValProSerGluArgProProProProValArgAspProGlyArgS 377
249 CA.....CCGAGGGGCAACCGGCTTTTCGATCAGCTGAGACATCA 209
|||||
377 erGlyProLeuProProProProValSerArgAsnGlySerThrSer 393
208 GCGGGTGGGTCAACGACCCACCTGCGCCA.....GGTAGCGAC 168
|||||
394 ArgAlaLeuProAlaThr..ProGlnLeuProSerArgSerGlyValAsp 409
167 TCCGCGCGCAGCAGCGCGCGCGCTGGGGCCTGATCCACGACCCAG 118
|||||
410 SerProArgSerGlyProArgProProLeuProProAspArgPro.... 424
117 CGGATGGTTCGACAGCGGACTGGTGCCGAGCAGGCCCATCTGCGGGCTT 68
424 ..... 424
67 CCTCGTGGGTGGTGGCGCGCGGTGGCGCCACC 30
|||||
425 .....SerAlaGlyAlaProProProProSer 435
```

OM of: US-09-462-480-3 to: Issued\_Patents\_AA:\* out\_format : pfs  
 Date: Jul 22, 2002 1:24 AM  
 About: Results were produced by the GenCore software, version 4.5,  
 Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame2\_n2p.model -DEV=xlh  
 -O=/cgn2\_1/USPTO\_spool/US09462480/runat\_18072002\_164418\_19425/app\_query.fasta\_1.2850  
 -DB-Issued\_Patents\_AA -OPWT-fastan -SUFFIX=rai -GAPOP=12.000  
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -DOOPEXT=0.000  
 -GAPOP=4.500 -GAPEXT=0.050 -YGAPOP=10.000 -YGAPEXT=0.500  
 -YGAPOP=6.000 -YGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsum62  
 -TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR\_SCORE=pct  
 -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFWT=pfs  
 -NORM-ext -HEARSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
 -USER=US09462480@cgn1\_1.18 -NCPU=6 -ICPU=3 -LONLOG  
 -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-462-480-3  
 Query length: 481  
 Database: Issued\_Patents\_AA\*  
 Database sequences: 231628  
 Database length: 24425594  
 Search time (sec): 53.850000

score\_list:

Sequence	Strd Orig	ZScore	Escore	Len	Documentation
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-08-818-112-114	..	713.00	1066.91	1.7e-52	368
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-08-818-111-109	+	713.00	1066.91	1.7e-52	368
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-09-056-536-114	+	713.00	1066.91	1.7e-52	368
/cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-09-010-928B-28	+	128.00	178.33	0.0026	745
/cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-09-010-928B-2	+	128.00	176.91	0.0026	870
/cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-08-726-306A-56	+	125.50	187.54	0.0032	181
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-09-413-814-78	-	125.00	172.26	0.0047	882
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-09-232-468A-8	-	121.00	173.39	0.0089	404
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-09-080-897-2	-	120.50	162.30	0.0119	1248
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-09-323-735-2	-	120.50	162.30	0.0119	1248
/cgn2_6/ptodata/2/iaa/6A_COMB.ppt:US-08-899-595-3	-	120.00	161.06	0.0133	1315
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-08-483-533-41	-	118.50	170.80	0.0141	355
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-09-283-471A-41	-	118.50	170.80	0.0141	355
/cgn2_6/ptodata/2/iaa/PCTUS_COMB.ppt:PCT-US91-06532-3	-	118.50	170.80	0.0141	355
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-09-010-928B-4	+	118.00	161.45	0.0183	907
/cgn2_6/ptodata/2/iaa/5A_COMB.ppt:US-07-945-283-2	+	117.50	153.64	0.0231	1958
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-08-483-533-40	-	116.00	169.75	0.0217	264
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-09-283-471A-40	-	116.00	169.75	0.0217	264
/cgn2_6/ptodata/2/iaa/5A_COMB.ppt:US-07-928-611-13	-	112.00	163.62	0.0472	267
/cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-08-487-811A-13	-	112.00	163.62	0.0472	267
/cgn2_6/ptodata/2/iaa/PCTUS_COMB.ppt:PCT-US93-07370-13	-	112.00	163.62	0.0472	267
/cgn2_6/ptodata/2/iaa/5A_COMB.ppt:US-08-056-051-6	-	112.00	158.49	0.0520	467
/cgn2_6/ptodata/2/iaa/5A_COMB.ppt:US-07-928-611-22	-	112.00	158.49	0.0520	467
/cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-08-487-811A-22	-	112.00	158.49	0.0520	467
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-09-060-694-22	-	112.00	158.49	0.0520	467
/cgn2_6/ptodata/2/iaa/PCTUS_COMB.ppt:PCT-US93-07370-22	-	111.50	158.49	0.0520	467
/cgn2_6/ptodata/2/iaa/PCTUS_COMB.ppt:PCT-US91-06532-2	-	111.00	163.00	0.0518	263
/cgn2_6/ptodata/2/iaa/backfilled.ppt:5168049-3	-	110.00	164.06	0.0650	183
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-08-483-533-26	-	110.00	160.91	0.0690	258
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-09-283-471A-26	-	110.00	160.91	0.0690	258
/cgn2_6/ptodata/2/iaa/5A_COMB.ppt:US-08-131-365B-38	-	110.00	147.24	0.0897	1147
/cgn2_6/ptodata/2/iaa/backfilled.ppt:5171843-9	+	109.50	156.66	0.0813	378
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-09-199-637A-351	-	109.00	154.37	0.0922	447
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-09-041-886-23	-	109.00	145.43	0.1094	1185
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-07-945-283-2	-	109.00	140.83	0.1195	1958
/cgn2_6/ptodata/2/iaa/backfilled.ppt:5171843-7	+	108.50	159.63	0.0905	232
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-08-556-978B-59	-	108.50	152.72	0.1033	493

/cgn2\_6/ptodata/2/iaa/6A\_COMB.ppt:US-08-586-165-4 + 108.00 154.87 0.1077 359  
 /cgn2\_6/ptodata/2/iaa/6B\_COMB.ppt:US-09-247-806-2 + 107.50 150.56 0.1270 529  
 /cgn2\_6/ptodata/2/iaa/5B\_COMB.ppt:US-08-728-323A-2 - 107.50 143.35 0.1458 1162  
 /cgn2\_6/ptodata/2/iaa/6B\_COMB.ppt:US-08-483-533-28 - 107.00 160.27 0.1144 169  
 /cgn2\_6/ptodata/2/iaa/6B\_COMB.ppt:US-09-283-471A-28 - 107.00 160.27 0.1144 169

seq\_name: /cgn2\_6/ptodata/2/iaa/6B\_COMB.ppt:US-08-818-112-114

seq\_documentation\_block:  
 ; Sequence 114, Application US/08818112  
 ; Patent No. 6290969  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Reed, Steven G.  
 ; APPLICANT: Skeiky, Yasir A.W.  
 ; APPLICANT: Dillon, Davin C.  
 ; APPLICANT: Campos-Neto, Antonio  
 ; APPLICANT: Houghton, Raymond  
 ; APPLICANT: Vedvick, Thomas S.  
 ; APPLICANT: Twardzik, Daniel R.  
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY  
 ; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS  
 ; NUMBER OF SEQUENCES: 153  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: SEED and BERRY LLP  
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue  
 ; CITY: Seattle  
 ; STATE: Washington  
 ; COUNTRY: USA  
 ; ZIP: 98104-7092  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/818,112  
 ; FILING DATE: 13-MAR-1997  
 ; CLASSIFICATION: 424  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Maki, David J.  
 ; REGISTRATION NUMBER: 31,392  
 ; REFERENCE/DOCKET NUMBER: 210121.411C6  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (206) 622-4900  
 ; TELEFAX: (206) 682-6031  
 ; INFORMATION FOR SEQ ID NO: 114:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 368 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; US-08-818-112-114

alignment\_scores:  
 Quality: 713.00 Length: 143  
 Ratio: 5.057 Gaps: 0  
 Percent Similarity: 98.601 Percent Identity: 96.503

alignment\_block:  
 US-09-462-480-3 x US-08-818-112-114 ..

Align seg 1/1 to: US-08-818-112-114 from: 1 to: 368

1 CTGACAGAGTGACGTCTTTTTCAGCCAGTGGCGGCGCCGCGGGG 50  
 |||||

226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGly 242

51 CAACCCAGCCGACGAGGAGCGCGAGATGGCTGCTCGGCACGATC 100

242 YasnProAlaaspGluGluAlaAlaGlnMetGlyLeuLeuGlyThrSerP 259

101 CGCTGTGGAACCATCCGCTGCTGTGTGATCAGGCCCGCGCGGCGCG 150

```
|||||
259 roLeuSerAsnHisProLeuAlaGlySerGlyProSerAlaGlyAla 275
151 GGCCTGCTCGCGCGGAGTCTGCTACCTGGCGCAGGTGGTGGTTCACCCG 200
|||||
276 GlyLeuLeuArgAlaGlySerLeuProGlyAlaGlySerLeuThrAr 292
|||||
201 CACGCCGCTGATGCTCAGCTGATCGAAAGCCGTTCCCCCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuLeuGluLysProValAlaProSerValM 309
|||||
251 TCCCGCGCGCTGTTCCCGATCGCTGACGGTGGGTCGCGCTCCGCTG 300
|||||
309 etProAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
|||||
301 GGTCCGGGAGCGATCGCGCAGGTTCGCAATCCCGCGCTCCACGAGCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
|||||
351 GGGTCTGTGCGCGCGGACCGCTCGCGCAGGAGCGTGAAGAAGACGACG 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAspG 359
|||||
401 AGGACGACTGGGACGAGGAGGAGTGG 429
|||||
359 luAspAspTrpAspGluGluAspAspTrp 368
```

seq\_name: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:US-08-818-111-109

```
seq_documentation_block:
; Sequence 109, Application US/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,111
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 109:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 368 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-818-111-109
```

```
alignment_scores:
  Quality: 713.00      Length: 143
  Ratio: 5.057         Gaps: 0
  Percent Similarity: 98.601  Percent Identity: 96.503

alignment_block:
US-09-462-480-3 x US-08-818-111-109 ..
Align seg 1/1 to: US-08-818-111-109 from: 1 to: 368
```

```
1 CTGACGAGGTGACGTCTGTTTTCAGCCAGGTGGCGCGCACCGCGCGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGlyGl 242
|||||
51 CAACCCAGCCGACGAGGAGGCGCGCAGATGGCTGCTCGGCACACAGT 100
|||||
242 yAsnProAlaAspGluGluAlaGlnMetGlyLeuLeuGlyThrSerP 259
|||||
101 CGCTGTGGAACCATCCGCTGGCTGGTGGATCAGGCCCGCGCGGCGCG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275
|||||
151 GGCCTGCTGCGCGCGGAGTCTGCTACCTGGCGCAGGTGGTGGTTCAC 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThr 292
|||||
201 CACGCCGCTGATGCTCAGCTGATCGAAAGCCGTTCCCGCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuLeuGluLysProValAlaProSerValM 309
|||||
251 TCCCGCGCGCTGTTCCCGATCGCTGACGGTGGGTCGCGCTCCGCTG 300
|||||
309 etProAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
|||||
301 GGTCCGGGAGCGATCGCGCAGGTTCGCAATCCCGCGCTCCACGAGCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
|||||
351 GGGTCTGTGCGCGCGGACCGCTCGCGCAGGAGCGTGAAGAAGACGACG 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAspG 359
|||||
401 AGGACGACTGGGACGAGGAGGAGTGG 429
|||||
359 luAspAspTrpAspGluGluAspAspTrp 368
```

seq\_name: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:US-09-056-556-114

```
seq_documentation_block:
; Sequence 114, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
```

```
;
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
;   NAME: Maki, David J.
;   REGISTRATION NUMBER: 31,392
;   REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (206) 622-4900
;   TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 114:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 368 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
; US-09-056-556-114
;
; alignment_scores:
;   Quality: 713.00      Length: 143
;   Ratio: 5.057        Gaps: 0
; Percent Similarity: 98.601 Percent Identity: 96.503
;
; alignment_block:
; US-09-462-480-3 x US-09-056-556-114
;
; Align seg 1/1 to: US-09-056-556-114 from: 1 to: 368
;
; 1 CTCGACGAGTGGTGTTCAGCGAGTGGCGGCGCGCGCGG 50
; |||||
; 226 LeuGlnValThrSerLeuPheSerGlnValGlyThrGlyGly 242
;
; 51 CAACCCAGCGAGGAGGAGCGCGGAGATGGGCTCTCGGCACCAAGTC 100
; |||||
; 242 YasnProAlaaspGluGluAlaGlnMetGlyLeuLeuGlyThrSerP 259
;
; 101 CGCTGTGCAACCATCCCTCGCTGGTGGATCAGGCCCGCGGCGCG 150
; |||||
; 259 roLeuSerAsnHisProLeuAlaGlySerGlyProSerAlaGlyAla 275
;
; 151 GGCTGTGCGCGGAGTGCCTACCTGCGGAGTGGTGGTGTGACCCG 200
; |||||
; 276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThr 292
;
; 201 CAGCGCGCTGATGTCTCAGCTGATCGAAAGCGGTTGCCCGCTCGGTGA 250
; |||||
; 292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerVal 309
;
; 251 TGCGCGCGCTGTTCGCGGATCGCTGACGGGTGGCGCGCTCCGGTG 300
; |||||
; 309 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
;
; 301 GGTCCGGGAGCGATGGCCAGGTTCCCAATCCGCGCGCTCCACCGCC 350
; |||||
; 326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
;
; 351 GGGTCTGTGCGCGCGCACCGTTCGCGGAGGAGCGTGAAGAAGACGACG 400
; |||||
; 342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluAspaspG 359
;
; 401 AGGACGACTGGGAGGAGGAGGACTGG 429
; |||||
; 359 luAspAspTrpAspGluAspAspTrp 368
```

seq\_name: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep:US-09-010-928B-28

```
seq_documentation_block:
; Sequence 28, Application US/09010928B
; Patent No. 5994099
; GENERAL INFORMATION:
; APPLICANT: Lewis, Randolph V
; APPLICANT: Hayashi, Cheryl Y
; TITLE OF INVENTION: EXTREMELY ELASTIC SPIDER SILK PROTEIN AND DNA
; TITLE OF INVENTION: CODING THEREFOR
```

```
;
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
; STREET: 8110 GATEHOUSE RD. SUITE 500E
; CITY: FALLS CHURCH
; STATE: VIRGINIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 22042
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/010,928B
; FILING DATE: 22-JAN-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M
; REGISTRATION NUMBER: 28977
; REFERENCE/DOCKET NUMBER: 1447-109P
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 745 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-010-928B-28
```

alignment\_scores:  
Quality: 128.00 Length: 132  
Ratio: 1.707 Gaps: 6  
Percent Similarity: 56.818 Percent Identity: 35.606

alignment\_block:

```
US-09-462-480-3 x US-09-010-928B-28
;
; Align seg 1/1 to: US-09-010-928B-28 from: 1 to: 745
;
; 37 GGCACCGCGGCGGCAACCCAGCCGACGAGGAGGAGCGGCGGCGCT 86
; ||| |||||
; 251 GlyProGlyGlyAlaGlyProGlyGlyAlaGlyProGlyGlyAlaGlyPr 267
;
; 87 GCTCGGCGACGAGTCCGCTGCGAACCATCCGCTGGTGGTGATCAGGCC 136
; ||| |||||
; 267 oGlyGlyTyrglyProGlyGlySerGlyPro.....GlyGlyAlaGlyP 282
;
; 137 CCAGCGGCGGCGCGGCGCTGCTGCGCGGAGTGCCTACCTGGCGCAGGT 186
; |||| |||||
; 282 roser...GlyAlaGlyLeuGlyGlyAlaGly.....ProGlyGlyAla 295
;
; 187 GGGTGTGTGACCCGACGCGCTGATGTCTCAGCTGATCGAAGACCGCGT 236
; |||
; 296 Gly.....
;
; 237 TGCCCCCTCGGTGATCGCGGCGCTGTTCGCGATCGTCTGCGTACGGGTG 286
; |||||
; 297 .....LeuGlyGlyAlaGlyProGlyGlyAlaGlyThrSerG 309
;
; 287 GCGCGCTCCG.....GTGGGTCCGGAGCGATGGGCCAGGTTCCGCAA 330
; |||||
; 309 lyAlaGlyProGlyGlyAlaGlyProGlyGlyAlaGlyGlnGlyAspAla 325
;
; 331 TCCGGCGGCTCCACCGCCC.....GGTCTGGTGGCGCGCGCGCGCTCG 376
; |||||
; 326 GlyProGlyGlyAlaGlyArgGlyGlyAlaGlyArgGlyGlyValGlyAr 342
;
; 377 GCGAGGAGCGGTGAAGAAGACGACGAGGAGGAGTGGGAGGAGGAGA 422
; |||||
; 342 gGlyGlyAlaGlyArgGlyGlyAlaGlyArgGlyGlyAlaArgGly 357
;
; seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-09-010-928B-2
```

[illegible]

36 aAlaAlaAlaGlyGlyProLeuArgTyGlyArgAlaValArgAlaArgG 53  
52 AACCC.....AGCCGAGGAGAGCCGCGCAGATGGCGCTGTCGCAC 95  
53 lyProArgAspAlaArgGlyAla.....AlaProGlyArgGly 66  
96 CAGTCCGCTCTGCAACCATCCGCTGGCTGGTGGATCATGCCGCCAGCGCGG 145  
67 AlaArgSerAlaThrProGlyAlaGly.....AlaProAlaArgG 80  
146 G.....CGGGCCCTGCTGCGCGCG 165  
80 yHisArgAlaArgAlaProAlaProArgArgGlyProAlaAlaArgG 97  
166 GAGTCGCTACCTGCGCGCAGGTGGTCTGTGACCGCAGCGCGCTGATGTC 215  
97 ly..... 97  
216 TCAGCTGATCGAAAGCGGTGTCCTCGGTGATCGCGCGGCTGTTG 265  
98 .....GlyArgGlyGlyProArgAlaGlyAlaLe 108  
266 CCGGATCGTGGTGACGGTGGCGCC.....GCTCCGGTGGTCCG 306  
108 uArgAlaGlyGlyArgGlyGlyAlaArgGlyProAlaGluGlyAla 124  
307 GGAGCGATGGCCAGGCTTCGCAATCCGCGGCTCCACCGCGCGG... 353  
125 GlyAlaAlaGlyGlyValArgLeuProAlaAlaProPro.ProGlyArg 140  
354 TCTGTGTCGCGCGCGCTCGCGCAGGAGCGTGAAGAAGACGACGAGG 403  
141 GlyGlyArgAlaAlaArgProAspProGlyLeuArgArgAlaGlyAl 157  
404 ACGACTGGGAGGAGAGGAGGAGTGGTGGTGGTCCCGTAATGAC 446  
157 aAspAlaGlyArgAspAlaArgArgProGluValArgArgAsp 171  
seq\_name: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep:US-09-413-814-78

seq\_documentation\_block:  
; Sequence 78, Application US/09413814  
; Patent No. 6225064  
; GENERAL INFORMATION:  
; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH  
; APPLICANT: Bristol-Myers Squibb, Co.  
; APPLICANT: Beyer, Stefan  
; APPLICANT: Bioecker, Helmut  
; APPLICANT: Brandt, Petra  
; APPLICANT: Cino, Paul M  
; APPLICANT: Dougherty, Brian A  
; APPLICANT: Goldberg, Steven L  
; APPLICANT: Hofle, Gerhard  
; APPLICANT: Mueller, Joachim  
; APPLICANT: Reichenbach, Hans  
; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or  
; TITLE OF INVENTION: heteropolyketide compounds  
; FILE REFERENCE: PCT/US 99/23535  
; CURRENT APPLICATION NUMBER: US/09/413,814  
; CURRENT FILING DATE: 1999-10-07  
; EARLIER APPLICATION NUMBER: DE 198 46 493.2  
; EARLIER FILING DATE: 1998-10-09  
; NUMBER OF SEQ ID NOS: 107  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 78  
; LENGTH: 882  
; TYPE: PRT  
; ORGANISM: Sorangium cellulosum  
US-09-413-814-78

alignment\_scores:  
Quality: 125.00 Length: 138

Ratio: 1.786 Gaps: 10  
Percent Similarity: 50.725 Percent Identity: 37.681  
alignment\_block:  
US-09-462-480-3/rev x US-09-413-814-78 ..  
Align seg 1/1 to: US-09-413-814-78 from: 1 to: 882  
424 GTCCTCTTCTGTCCTCCAGTCCTCTCTCTTTCACGCTCCTCGG... 377  
347 ArgProThrArgProSerArgProArgArg.....ThrProSerAr 360  
376 ....CGAGCGGTGCGCGCGGACAGACCGCGGCTGTGGAGCGCGCGGA 331  
360 gArgArgArgSerProAlaArg..... 367  
330 TTGGGAACCTTGGCCATCGTCCCGGACCCACCGGAGCGGCGCCACCG 281  
368 ....ThrProGlyProArgSerPro.....ArgArgArgPro 378  
280 TCACCGGACGATCCGG.....CAACAGCGCGCGGCATCACC 246  
379 SerProSerAlaArgSerProAlaIleTrpGluArgProArgArgPr 395  
245 GAGGGGCAACCGGCTTTTCATCAGCTGAGACATCAGCGCGTGGTGGT 196  
395 oArgAspArgArgProArgArg.LeuLeuArgProAlaArgArgLeuAla 411  
195 CAACGACCCACCT.....GCGCCAGTAGGACTCCGCGCGGACGAGGC 152  
412 ArgArg.AspproGlyAspValProAspProArgAlaAlaArgArg. 427  
151 CCGCGCGCGCTGGCGCTGATCCACCGCGGATGTTGTCGACAGC 102  
428 ..AlaProAlaAlaGlyAlaLeuProGlyAlaAspGlyArgGlyAlaArg 443  
101 GGACTGTGTGCGGAGCGGCGCATCTGCGCGGCTTCTCTCGTGGGT 52  
444 Gly.....ProGlyArgArgArgAlaProArgArgGlyAlaAlaAr 458  
51 GCGCGCGCGCG 42  
458 gValProPro 461  
seq\_name: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep:US-09-232-468A-8

seq\_documentation\_block:  
; Sequence 8, Application US/09232468A  
; Patent No. 6207165  
; GENERAL INFORMATION:  
; APPLICANT: AUDONNET et al.  
; TITLE OF INVENTION: POLYNUCLEOTIDE VACCINE FORMULA AGAINST PORCINE  
; TITLE OF INVENTION: REPRODUCTIVE AND RESPIRATORY PATHOLOGIES  
; FILE REFERENCE: 454313-2230  
; CURRENT APPLICATION NUMBER: US/09/232,468A  
; CURRENT FILING DATE: 1999-01-05  
; NUMBER OF SEQ ID NOS: 54  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 8  
; LENGTH: 404  
; TYPE: PRT  
; ORGANISM: Aujeszky's Disease Virus (NIA3 Strain)  
US-09-232-468A-8

alignment\_scores:  
Quality: 121.00 Length: 166  
Ratio: 1.424 Gaps: 7  
Percent Similarity: 51.205 Percent Identity: 30.723

alignment\_block:  
US-09-462-480-3/rev x US-09-232-468A-8 ..

Align seg 1/1 to: US-09-232-468A-8 from: 1 to: 404

472 CGGCGCGGTGCGCGGAGTCTGTTCATACGGAG...CTCACCAG 426  
||||| |||: |||: |||: |||: |||  
193 ArgProArgGlyProAlaProHisValGlnValArgValLeuGluArg 209  
425 TCGTCTCTCTGCTCCAGTCGTCGCTCTTTCACGCTCCTCGGC 376  
||| |||: |||: |||: |||: |||  
209 gargGlnLeuGlnAlaGlyArgGlyArgAspAlaile.....ProAspA 224  
375 GAGCGGTGCGCGGACGACACCGCGGTGTGGAGCGCGCGGATTCG 326  
: |||: |||: |||: |||: |||  
224 laValLeuProAlaProAlaProAlaProGlyGlyGlyLeuLeuValPro 240  
325 AACCTGCGCATCTGCTCCGCGACCCACCGAGCGCGCGCCACCGCTACC 276  
: |||: |||: |||: |||: |||  
241 GlnGluArgProAspAlaProAlaGlyProArgArgHis..... 254  
275 GACGATCGGCAACACCGCGCGCA.....TCAC 247  
: |||: |||: |||: |||: |||  
255 .AlaValArgHisArgProArgAlaAlaLeuGlyGlyLeuAlaGluAlap 271  
246 CGAGGGGCAACCGCTTTTCATCAGCTGAGACATCAGCGCGGTGGCGG 197  
|| ||| |||: |||: |||: |||  
271 roAlaProAlaProAlaProAlaProAlaGluAlaArgAlaArgProGly 287  
196 TCAAGACCCACCTGCGCGAGGTAGCAGCTCCGCGCGCAGCGCGCG 147  
: |||: |||: |||: |||: |||  
288 AspAlaArgAlaProArgProProAlaArgAlaGlyAspAlaGlyProAr 304  
146 CCCGCTGGGCGCTGATCCACCGCGCGGTGTCGACAGCGGACT 97  
| ||| |||: |||: |||: |||  
304 gArgArg.....GlyProProHisAlaAlaThrProGluA 316  
96 GGTGCGCGCAGCGC.....CCATCTGCGCGCTTCTCTCGTCGG 59  
|||: |||: |||: |||: |||  
316 laArgAspAlaAlaProProLeuArgProAlaGlyArgAlaGlnArg 332  
58 CTGGTTCGCCCGCGGTGCGCGCAGCTGCTGAGAACAGCAGT 13  
: |||: |||: |||: |||: |||  
333 ValAlaAlaAlaArgGlyAlaValProAlaAlaAlaAspProArg 347

seq\_name: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:US-09-080-897-2

seq\_documentation\_block:

; Sequence 2, Application US/09080897

; Patent No. 5985574

; GENERAL INFORMATION:

; APPLICANT: King, Mary-Claire

; APPLICANT: Lynch, Eric D.

; APPLICANT: Lee, Ming

; APPLICANT: Morrow, Jan E.

; APPLICANT: Welcsh, Piri L.

; APPLICANT: Leon, Pedro E.

; TITLE OF INVENTION: Modulators of Actin

; NUMBER OF SEQUENCES: 14

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP

; STREET: 75 DENISE DRIVE

; CITY: HILLSBOROUGH

; STATE: CALIFORNIA

; COUNTRY: USA

; ZIP: 94010

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/080,897

; FILING DATE:

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: OSMAN, RICHARD A  
; REGISTRATION NUMBER: 36,627  
; REFERENCE/DOCKET NUMBER: UW97-001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650) 343-4341  
; TELEFAX: (650) 343-4342  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1248 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-09-080-897-2

alignment\_scores:  
Quality: 120.50 Length: 149  
Ratio: 1.826 Gaps: 8  
Percent Similarity: 44.295 Percent Identity: 33.557

alignment\_block:  
US-09-462-480-3/rev x US-09-080-897-2

Align seg 1/1 to: US-09-080-897-2 from: 1 to: 1248

448 TTGTCAATTACGGAGCTCACCAGTCGTCTCTTCTGTCCTCCAGTCGTCCTCG 399  
||||| |||: |||: |||: |||: |||  
556 LeuSerAlaAlaAlaThrValProProSerValProSerArg..... 570  
398 TCGTCTCTTTCACGCTCCTGCGAGCGGTGCGCGCGGACGACGACCGCG 349  
||| |||: |||: |||: |||: |||  
571 .....AlaProValProProAlaProLeuProG 581  
348 GCTGTGTGAGCGCGCGGATTGCGAACCTGCGCCATCGCTCCCGGACCCA 299  
|| |||: |||: |||: |||: |||  
581 ly.....AspSerGlyThrIleIleProProPro 590  
298 CCG.....GAGCGCGCGCACCGTCACCGAGCATCCGGCAACAGCC 258  
||| |||: |||: |||: |||: |||  
591 ProAlaProGlyAspSerThrThrProProProProProProProProPr 607  
257 GCGGCGATCACCAG.....GGGCAACCGGCTTTTCGATCAGTCAGTGAGACA 212  
|||| |||: |||: |||: |||: |||  
607 oProProProLeuProGlyGlyThrAlaIleSerProProProProL 624  
211 TCAGCGCGTGGGTCAACGACCGCATCGTCCGCGCA..... 177  
: |||: |||: |||: |||: |||  
624 euSerGlyAspAlaThrIleProProProProProProLeuProGluGlyVal 640  
176 GGTAGCGACTCCGCGCGCAGCAGGCGCC...CGCGCGCGGTGGGGCCTGA 130  
||| |||: |||: |||: |||: |||  
641 GlyIleProSerProSerSerLeuProGlyGlyThrAlaIleProProPr 657  
129 TCCACCGACCGCGGATGTTTCGACGCGGACTGTCGCGGACGCGCGCA 80  
|||| |||: |||: |||: |||: |||  
657 oProProLeuProGly.....SerAlaArgIleProProProPro 670  
79 TCTGCGCGGTTCCTGCTGCGTGGTTCGCGCGCGGTGCGCGCGCC 33  
: |||: |||: |||: |||: |||  
671 ..ProProLeuProGlySerAlaGlyIleProProProProProProPro 685

seq\_name: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:US-09-323-735-2

seq\_documentation\_block:

; Sequence 2, Application US/09323735

; Patent No. 6197932

; GENERAL INFORMATION:

; APPLICANT: King, Mary-Claire

; APPLICANT: Lynch, Eric D.

; APPLICANT: Lee, Ming

; APPLICANT: Morrow, Jan E.

; APPLICANT: Welcsh, Piri L.

; APPLICANT: Leon, Pedro E.





398 TCGTCTCTTTCACGGCTCT ..... 380  
623 ..... AlaProValProAlaProProLeuProGlyAspS 635  
379 ..GCGGAGCGGTGCGCGGACACCGCGGTGTGTGGAGCGCGG 332  
635 erGlyThrIleProProAlaProGly ..... AspSer 648  
331 ATTGGACCCCTGGCCCTCGCTCCGACCCAGCGGAGCGGCCACCC 282  
649 ThrThrProProProProProProProProProProProPro 665  
281 GTCACCGACGATCCGCGCACCGCGCATCCAGAGGGGCAACCGG 232  
665 oGlyGlyValCysIleSerProProSerLeuPro..GlyGlyThrAl 681  
231 CTTTTCATCAGCTGAGATCAGCGCGGTGCGGTCAACACCCACCTG 182  
681 aileSerProProProProLeuSerGlyAspAlaThrIleProPro 698  
181 CGCCA ..... GGTACGACTCGCGCGCAGCAGGCC... 150  
698 roProLeuProGluGlyValGlyIleProSerProSerLeuProGly 714  
149 GCGCGCGCGTGGCGCTGATCCACCGCACCGGATGTTCCAGACGGG 100  
715 GlyThrAlaIleProProProProProProGly ..... SerAl 728  
99 ACTGTGCGGAGCAGCGCCATCTGCGCGGTCTCTGTCGGGTGGTGC 50  
728 aArgIleProProProPro...ProProLeuProGlySerAlaGlyIleP 744  
49 CGCGCGCGTGGCGGCC 33  
744 roProProProPro 749  
seq\_name: /cgn2\_5/podata/2/iaa/6B\_COMB.pep:US-08-483-533-41  
seq\_documentation\_block:  
; Sequence 41, Application US/08483533  
; Patent No: 6172047  
; GENERAL INFORMATION:  
; APPLICANT: Roizman, Bernard  
; APPLICANT: Chou, Joany  
; TITLE OF INVENTION: Method for Treating Tumorigenic  
; TITLE OF INVENTION: Diseases  
; NUMBER OF SEQUENCES: 43  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Bdrun  
; STREET: 6300 Sears Tower, 233 South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States of America  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/483,533  
; FILING DATE: 07-MAR-95  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/419,853  
; FILING DATE: 11-APR-95  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/861,233  
; FILING DATE: 31-MAR-92  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Zeller, James P.  
; REGISTRATION NUMBER: 28,491

; REFERENCE/DOCKET NUMBER: 28097/32742  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312/474-6300  
; TELEFAX: 312/474-0448  
; TELEX: 25-3856

; INFORMATION FOR SEQ ID NO: 41:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 355 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-483-533-41

alignment\_scores:  
Quality: 118.50 Length: 222  
Ratio: 1.411 Gaps: 9  
Percent Similarity: 37.838 Percent Identity: 26.126

alignment\_block:

US-09-462-480-3/rev x US-08-483-533-41 ..

Align seg 1/1 to: US-08-483-533-41 from: 1 to: 355

440 ACGGAGCT.....CACCAGTCGTCTCTTCGTCGCCAGTCGTCTC 400  
|||||  
19 ThrGlyAlaValProThrAlaGlnSerGlnValThrSerThrProAsnSe 35

399 GTGCTCTTTCACGCTCTCGCGGAGCGGTGCGCGGACGACGACCGG 350  
|||  
35 rGluProAlaValArgSerAlaProAlaAlaProProProProA 52

349 GGCTGTGGAGCCCGCGGATTGC..... 327  
|||  
52 laGlyGlyProProProSerCysSerLeuLeuLeuArgGlnTrpLeuHis 68

336 .....GAACCTGTGCCCATCGC 310  
|||  
69 ValProGluSerAlaSerAspAspAspAspAspTrpProAspSe 85

309 TCCC.....GGACCCACGAGCGCGG.... 288  
|||  
85 rProProGluSerAlaProGluAlaArgProThrAlaAlaProA 102

287 ..CCACCGTCCACGACGATCGGCA.....ACA 261  
|||||  
102 rProProGlyProHisArgProAlaTrpAlaArgGlyAlaGlyLeuThr 118

260 GCGCGCGCATCACGAGGGGCAACCGCTTTCGATCAGCTGAGACAT 211  
|||||  
119 .ProProThrProProArgAlaProSerAlaPheArgAlaSerProS 135

210 CAGCGCGTGGGTCAACGACCCACCTGCGCGGACGAGTACGACTCCGCG 161  
|||||  
135 erAlaCysAlaSerProArgSerThrTrpArgAlaCysAlaCysAspAla 151

160 GCACGAGC..... 152  
|||||  
152 ArgAlaGlyArgGlyArgArgSerProProArgProProArgProProAr 168

151 .....CGCGCGCGCGTGGGCGCTGATCCACCGACCGCGG 115  
|||||  
168 gProProArgProProArgProProArgGlyCysAlaSerArgProThrS 185

114 ATGGTTCGACAGCGGACTGTGCGCGGACGAGCCCATCTGCG..... 74  
|||  
185 erGlyCysAlaThrTrpSerGlyProArgProProAlaTrpArgAla 201

74 ..... 74  
202 AlaAlaArgGlyProAlaSerGlyProThrGlyLeuGlySerGlyAlaGl 218



```

114 ATGGTTCGACAGCGGAAGCTGTCCGCAGCAGGCCCATCTGCG.....74
    :||| :|||| ||| ||| |||
185 egiGysAlaThrTrpSerGlyProArgProAlaTrpArgAla 201
74 .....74
202 AlaAlaArgGlyProAlaSerGlyProThrGlyLeuGlySerGlyAlaGl 218
73 ....CGGTTCTCTCGTGCGGTGGGTG.....CGCGCGCCGG 41
    ||| ||||| ||| :||| |||
218 yTrpArgAtgProArgArgSerGlyArgAlaTrpGlyProArgProv 235
40 TGC CGCCCACTGG 27
    |||| |
235 alProGlyProTrp 239
seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB pep:US-09-010-928B-4

seq_documentation_block:
; Sequence 4, Application US/09010928B
; Patent No. 5994099
; GENERAL INFORMATION:
; APPLICANT: Lewis, Randolph V
; APPLICANT: Hayashi, Cheryl Y
; TITLE OF INVENTION: EXTREMELY ELASTIC SPIDER SILK PROTEIN
; TITLE OF INVENTION: CODING THEREFOR
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
; STREET: 8110 GATEHOUSE RD. SUITE 500E
; CITY: FALLS CHURCH
; STATE: VIRGINIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 22042
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/010,928B
; FILING DATE: 22-Jan-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M
; REGISTRATION NUMBER: 28977
; REFERENCE/DOCKET NUMBER: 1447-109P
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 907 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-010-928B-4

alignment_scores:
Quality: 118.00 Length: 119
Ratio: 1.662 Gaps: 6
Percent Similarity: 59.664 Percent Identity: 35.294

alignment_block:
US-09-462-480-3 x US-09-010-928B-4 ...

Align seg 1/1 to: US-09-010-928B-4 from: 1 to: 907

37 GCACCGCGCGGCAACCACCGCACGAGGAGAACCGCGCAGATGGCGCT 86
   ||| ||||| :||| :||| :|||
186 GlyProGlyGlyAlaGlyProGlyGlyAlaGlyProGlyGlyAlaGlyPr 202
      :||| :||| :||| :||| :|||
87 GCTCGGCAACCACTCCGCTGTCCGAACCACTCCGCTGGCTGGTGGATCAGGCC 136
   ||| :||| :||| :||| :|||

```

202 oGlyGlyAlaGlyProGlyGlyAlaGlyPro.....GlyGlyAlaGlyG 217  
137 CCAGCGCGGCGCGGCGCTGCTGCGCGGAGTCGCTACCTGGCGCAGGT 186  
: : : : : ||||| ||| : : : |||||  
217 LyAlaGlyGlyAlaGly.....GlyAlaGlyGlySerGlyGlyAlaGly 231  
187 GGGTCGTGACCGCGCAGCGCGCTGATGTCACGTG..... 222  
||||| ||| : : : : : |||||  
232 GlySer...GlyGlyThrThrIleIleGluAspLeuAspIleThrIleAs 247  
223 ....ATCGAAAAGCGGTTGCCCCCTCG...GTGATGCCGCGGCTGTG 265  
: : : : : ||||| ||| : : : : :  
247 pGlyAlaAspGlyProIleThrIleSerGluGluLeuThrIleSerGlyA 264  
266 CCGATTCGTCGGTGACGGTGCGCGCTCCG.....GTGGTCCGGGA 309  
||| : : : ||| ||||| : : : ||| |||||  
264 IaGlyGlySerGlyProGlyGlyAlaGlyProGlyGlyValGlyProGly 280  
310 CGGATGGCGCGGTTCCGAATCCGGCGGCTCCACCGCGGTCGTGTGT 359  
: : : ||| : : : : : ||||| : : : |||  
281 GlySerGlyProGlyGlyValGlyProGlyGlySerGlyProGlyGlyVa 297  
360 CGCACC 366  
297 IGlyPro 299

THIS PAGE BLANK (USPTO)

OM of: US-09-462-480-3 to: Pending\_Patents\_AA\_Main:\* out\_format : pfs  
Date: Jul 22, 2002 1:35 AM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:  
-MODEL=frame+n2p.model -DEV=xlh  
-O=/cgn2\_1/usptool/US09462480/runat\_18072002\_164419\_19468/app\_query.fasta\_1.2850  
-DB=Pending\_Patents\_AA\_Main -QFMT=fastan -SUFFIX=rapm  
-GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000  
-LOOPEXT=0.000 -OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000  
-XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000  
-YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1  
-MATRIX=blosom62 -TRANS=human40.cgi -LIST=45 -DALIGN=200  
-THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pfs -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09462480\_SCGN1\_1\_271 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

Search information block:  
Query: US-09-462-480-3  
Query length: 481  
Database: Pending\_Patents\_AA\_Main:\*  
Database sequences: 3502263  
Database length: 351980561  
Search time (sec): 542.380000

score\_list:

Sequence	Strd Orig	zScore	Escore	Len	Documentation
/cgn2_6/ptodata/2/paa/PCTUS_COMB.ppt:US99-03265-109	+	713.00	858.99	9.2e-40	368
/cgn2_6/ptodata/2/paa/PCTUS_COMB.ppt:US99-03265-114	+	713.00	858.99	9.2e-40	368
/cgn2_6/ptodata/2/paa/US086_COMB.ppt:US-08-658-800-109	+	713.00	858.99	9.2e-40	368
/cgn2_6/ptodata/2/paa/US086_COMB.ppt:US-08-659-683-114	+	713.00	858.99	9.2e-40	368
/cgn2_6/ptodata/2/paa/US086_COMB.ppt:US-08-680-573-109	+	713.00	858.99	9.2e-40	368
/cgn2_6/ptodata/2/paa/US086_COMB.ppt:US-08-680-574-114	+	713.00	858.99	9.2e-40	368
/cgn2_6/ptodata/2/paa/US087_COMB.ppt:US-08-729-622-109	+	713.00	858.99	9.2e-40	368
/cgn2_6/ptodata/2/paa/US087_COMB.ppt:US-08-730-510-114	+	713.00	858.99	9.2e-40	368
/cgn2_6/ptodata/2/paa/US089_COMB.ppt:US-08-942-341-109	+	713.00	858.99	9.2e-40	368
/cgn2_6/ptodata/2/paa/US089_COMB.ppt:US-08-942-578-114	+	713.00	858.99	9.2e-40	368
/cgn2_6/ptodata/2/paa/US090_COMB.ppt:US-09-024-753-109	+	713.00	858.99	9.2e-40	368
/cgn2_6/ptodata/2/paa/US090_COMB.ppt:US-09-025-197-114	+	713.00	858.99	9.2e-40	368
/cgn2_6/ptodata/2/paa/US090_COMB.ppt:US-09-072-596-109	+	713.00	858.99	9.2e-40	368
/cgn2_6/ptodata/2/paa/US090_COMB.ppt:US-09-072-967-114	+	713.00	858.99	9.2e-40	368
/cgn2_6/ptodata/2/paa/US097_COMB.ppt:US-09-724-685-114	+	713.00	858.99	9.2e-40	368
/cgn2_6/ptodata/2/paa/US092_COMB.ppt:US-09-050-739-92	+	713.00	858.92	9.2e-40	371
/cgn2_6/ptodata/2/paa/US092_COMB.ppt:US-09-246-191-92	+	713.00	858.92	9.2e-40	371
/cgn2_6/ptodata/2/paa/US098_COMB.ppt:US-09-791-171-92	+	713.00	858.92	9.2e-40	371
/cgn2_6/ptodata/2/paa/US099_COMB.ppt:US-09-804-980-92	+	713.00	858.92	9.2e-40	371
/cgn2_6/ptodata/2/paa/US090_COMB.ppt:US-09-943-443-90	-	373.00	457.79	5.4e-17	139
/cgn2_6/ptodata/2/paa/US097_COMB.ppt:US-09-791-171-90	-	373.00	457.79	5.4e-17	139
/cgn2_6/ptodata/2/paa/US098_COMB.ppt:US-09-804-980-90	-	373.00	457.79	5.4e-17	139
/cgn2_6/ptodata/2/paa/US099_COMB.ppt:US-09-943-443-90	-	373.00	457.79	5.4e-17	139
/cgn2_6/ptodata/2/paa/PCTUS_COMB.ppt:US00-30037-144	-	147.50	179.80	0.0734	314
/cgn2_6/ptodata/2/paa/PCTUS_COMB.ppt:US01-11988-592	-	147.50	179.80	0.0734	314
/cgn2_6/ptodata/2/paa/PCTUS_COMB.ppt:US02-08278-1613	-	147.50	179.80	0.0734	314
/cgn2_6/ptodata/2/paa/US098_COMB.ppt:US-09-833-245-592	-	147.50	179.80	0.0734	314
/cgn2_6/ptodata/2/paa/US092_COMB.ppt:US-09-252-991A-21297	-	146.00	174.56	0.0943	4
/cgn2_6/ptodata/2/paa/PCTUS_COMB.ppt:US00-05989-700	-	145.50	181.07	0.0979	200
/cgn2_6/ptodata/2/paa/US099_COMB.ppt:US-09-925-297-700	-	145.50	181.07	0.0979	200
/cgn2_6/ptodata/2/paa/US095_COMB.ppt:US-09-599-287A-2	-	145.50	173.56	0.1021	503
/cgn2_6/ptodata/2/paa/US100_COMB.ppt:US-10-078-547-2	-	145.50	173.56	0.1021	503
/cgn2_6/ptodata/2/paa/US092_COMB.ppt:US-09-252-991A-31950	-	143.50	179.66	0.1326	1
/cgn2_6/ptodata/2/paa/US092_COMB.ppt:US-09-252-991A-21881	-	142.00	166.42	0.1781	7
/cgn2_6/ptodata/2/paa/PCTUS_COMB.ppt:US01-04098A-3487	-	141.50	162.43	0.1961	1
/cgn2_6/ptodata/2/paa/US092_COMB.ppt:US-09-252-991A-17143	-	140.50	169.37	0.2187	4
/cgn2_6/ptodata/2/paa/US092_COMB.ppt:US-09-252-991A-30739	-	140.50	165.81	0.2230	6

/cgn2\_6/ptodata/2/paa/US095\_COMB.ppt:US-09-599-287A-24 - 140.00 166.87 0.2387  
/cgn2\_6/ptodata/2/paa/US100\_COMB.ppt:US-10-078-547-24 - 140.00 166.87 0.2387  
/cgn2\_6/ptodata/2/paa/US092\_COMB.ppt:US-09-252-991A-29050 - 139.50 168.60 0.2546  
/cgn2\_6/ptodata/2/paa/US092\_COMB.ppt:US-09-252-991A-16754 - 139.50 165.82 0.2585  
/cgn2\_6/ptodata/2/paa/US092\_COMB.ppt:US-09-252-991A-25333 - 139.00 169.81 0.2722

seq\_name: /cgn2\_6/ptodata/2/paa/PCTUS\_COMB.ppt:US99-03265-109

seq\_documentation\_block:  
; Sequence 109, Application PC/TUS9903265  
; GENERAL INFORMATION:  
; APPLICANT: Corixa Corporation  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF  
; TITLE OF INVENTION: TUBERCULOSIS  
; NUMBER OF SEQUENCES: 350  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds, LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036-2811  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US99/03265  
; FILING DATE: 17-FEB-1999  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 09/024,753  
; FILING DATE: 18-FEB-1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Laura A. Coruzzi  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 9532-0023-228  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-493-4935  
; TELEFAX: 650-493-5556  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 109:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 368 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; PCT-US99-03265-109

alignment\_scores:  
Quality: 713.00 Length: 143  
Ratio: 5.057 Gaps: 0  
Percent Similarity: 98.601 Percent Identity: 96.503

alignment\_block:  
US-09-462-480-3 x PCT-US99-03265-109 ..

Align seg 1/1 to: PCT-US99-03265-109 from: 1 to: 368

1 CTGCAGCAGGTGACGCTGTTTCAGGCAGGTGGCGGCACCGCGCGCGG 50  
|||||  
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGlyG 242  
51 CAACCCAGCCGACGAGGAGCCGCGCAGATGGCGCTGCTGGCACCAGTC 100  
|||||  
242 YAsnProAlaAspGluGluAlaAlaGlnMetGlyLeuLeuGlyThrSerP 259  
101 CGCTGTGACACCATCCCTGCTGGTGGATCAGGCCCGCGCGCGCGCGG 150  
|||||  
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275

```
151 GGCCTGCTGCGCGGAGTCGCTACCTCGCGCAGGTGGGTGTCGACCCG 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlySerLeuThrAr 292
|||||
201 CACCCCGCTGATGCTCAGCTGATCGAAAGCCGGTTCGCCCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309
|||||
251 TGCGGCGGCTGTTGCCGGATGCTGCGGTGACGGGTGGCGCGCTCCGGTG 300
|||||
309 etProAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
|||||
301 GGTCCGGAGGATGGCGCAGGTTCGCAATCCGGCGCTCCACAGCCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
|||||
351 GGGTCTGTCGCGCGGCAACGCTCGCGCAGGAGCGTGAAGAAGACGACG 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluAspAspG 359
|||||
401 AGGACGACTGGGACGAGGACGACTGG 429
|||||
359 luAspAspTipAspGluAspAspTrp 368
|||||
seq_name: /cgn2_5/ptodata/2/paa/US99-03268-114
seq_documentation_block:
; Sequence 114, Application PC/TUS9903268
; GENERAL INFORMATION:
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds, LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US99/03268
; FILING DATE: 17-FEB-1999
; CLASSIFICATION: 406
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/025,197.
; FILING DATE: 18-FEB-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 9532-0013-228
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-493-4935
; TELEFAX: 650-493-5556
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 114:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 368 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; PCT-US99-03268-114
```

```
alignment_scores:
  Quality: 713.00      Length: 143
  Ratio: 5.057        Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503
```

alignment\_block:

US-09-462-480-3 x PCT-US99-03268-114

Align seg 1/1 to: PCT-US99-03268-114 from: 1 to: 368

```
1 CTGACGAGGTGACGTCGTTTTCAGCCAGGTGGCGGCGCACCGGGCGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGlyG 242
|||||
51 CAACCCAGCGCAGCAGGAGGCGCGCAGATGGGCTGCTCGGCACGATC 100
|||||
242 yasnProAlaAspGluGluAlaAlaGlnMetGlyLeuLeuGlyThrSerP 259
|||||
101 CGCTGTCGAACCATCCGCTGGTGGTGGATCAGGCCCGCAGCGCGCGG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275
|||||
151 GGCTGCTGCGCGGGAGTCTGCTACCTGGCGCAGGTGGGTGCTTACCCG 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292
|||||
201 CACGCCGCTGATGCTCAGCTGATCGAAAGCCGGTTCGCCCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309
|||||
251 TCGCGCGGCTGTTGCCGGATGCTGCGGTGACGGGTGGCGCGCTCCGGTG 300
|||||
309 etProAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
|||||
301 GGTCCGGAGGATGGCGCAGGTTCGCAATCCGGCGCTCCACAGCCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
|||||
351 GGGTCTGTCGCGCGGCGCAGCTCGCGCAGGAGCGTGAAGAAGACGACG 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluAspAspG 359
|||||
401 AGGACGACTGGGACGAGGACGACTGG 429
|||||
359 luAspAspTipAspGluAspAspTrp 368
|||||
```

seq\_name: /cgn2\_5/ptodata/2/paa/US086\_COMB.pcp:US-08-658-800-109

seq\_documentation\_block:

; Sequence 109, Application US/08658800

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; APPLICANT: Skeiky, Yasir A.W.

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF

TUBERCULOSIS

; NUMBER OF SEQUENCES: 123

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEED and BERRY LLP

; STREET: 6300 Columbia Center, 701 Fifth Avenue

; CITY: Seattle

; STATE: Washington

; COUNTRY: USA

; ZIP: 98104-7092

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/658,800

; FILING DATE: 05-JUN-1996

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Maki, David J.

; REGISTRATION NUMBER: 31,392

; REFERENCE/DOCKET NUMBER: 210121.417C3

; TELECOMMUNICATION INFORMATION:



; TELEPHONE: (206) 622-4900  
 ; TELEFAX: (206) 682-6031  
 ; INFORMATION FOR SEQ ID NO: 109:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 368 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; US-08-658-800-109

alignment\_scores:  
 Quality: 713.00 Length: 143  
 Ratio: 5.057 Gaps: 0  
 Percent Similarity: 98.601 Percent Identity: 96.503

alignment\_block:  
 US-09-462-480-3 x US-08-658-800-109 ..

Align seg 1/1 to: US-08-658-800-109 from: 1 to: 368

1 CTCAGCAGGTGACGCTGTTGTTTCAGCCAGGTGGCGGCGCACCGCGCGG 50  
 |||||  
 226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGlyG 242  
 |||||  
 51 CAACCCAGCGGACGAGGAGGAGCGCGCAGATGGGCCTGCTCGGCACCAAGTC 100  
 |||||  
 242 YAsnProAlaAspGluGluAlaGlnMetGlyLeuLeuGlyThrSerP 259  
 |||||  
 101 CGCTGTCGAACCATCCGCTGGCTGGTGATCAGGCCCGCGCGCGCGGCGG 150  
 |||||  
 259 roLeuSerAsnHisProLeuAlaGlySerGlyProSerAlaGlyAla 275  
 |||||  
 151 GGCTGCTGCGCGCGGAGTGCTACCTACCTGGCGCAGGTGGCTGTGACCCG 200  
 |||||  
 276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292  
 |||||  
 201 CACGCCGCTGATGTCTCAGCTGATCGAAGAACCGGTTGCCCGCTCGGTGA 250  
 |||||  
 292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309  
 |||||  
 251 TGCCGGCGGCTGTTGCCGGATCGCTGGTGACGGGTGGCGCGCGCTCCGGTG 300  
 |||||  
 309 etProAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325  
 |||||  
 301 GGTCCGGGACGATGGCCAGGTTCCGAATCCGGCGGCTCCACCAAGCC 350  
 |||||  
 326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342  
 |||||  
 351 GGCTGCTGTCGCCCGCACCGCTCGCGCAGGAGCGGTGAAGAACGACGACG 400  
 |||||  
 342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAspG 359  
 |||||  
 401 AGGACGACTGGGACGAGAGGACTGG 429  
 |||||  
 359 luAspAspTrpAspGluGluAspAspTrp 368

seq\_name: /cgn2\_6/ptodata/2/paa/US086\_COMB.pap:US-08-659-683-114

seq\_documentation\_block:  
 ; Sequence 114, Application US/08659683  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Reed, Steven G.  
 ; APPLICANT: Skeiky, Yasir A.W.  
 ; APPLICANT: Dillon, Davin C.  
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY  
 ; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS  
 ; NUMBER OF SEQUENCES: 128  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: SEED AND BERRY LLP  
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue  
 ; CITY: Seattle  
 ; STATE: Washington

; COUNTRY: USA  
 ; ZIP: 98104-7092  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/659,683  
 ; FILING DATE: 05-JUN-1996  
 ; CLASSIFICATION:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Maki, David J.  
 ; REGISTRATION NUMBER: 31,392  
 ; REFERENCE/DOCKET NUMBER: 210121.411C3  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (206) 622-4900  
 ; TELEFAX: (206) 682-6031  
 ; INFORMATION FOR SEQ ID NO: 114:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 368 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; US-08-659-683-114

alignment\_scores:  
 Quality: 713.00 Length: 143  
 Ratio: 5.057 Gaps: 0  
 Percent Similarity: 98.601 Percent Identity: 96.503

alignment\_block:

US-09-462-480-3 x US-08-659-683-114 ..

Align seg 1/1 to: US-08-659-683-114 from: 1 to: 368

1 CTCAGCAGGTGACGCTGTTGTTTCAGCCAGGTGGCGGCGCACCGCGCGG 50  
 |||||  
 226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGlyG 242  
 |||||  
 51 CAACCCAGCGGACGAGGAGGAGCGCGCAGATGGGCCTGCTCGGCACCAAGTC 100  
 |||||  
 242 YAsnProAlaAspGluGluAlaGlnMetGlyLeuLeuGlyThrSerP 259  
 |||||  
 101 CGCTGTCGAACCATCCGCTGGCTGGTGATCAGGCCCGCGCGCGCGCGG 150  
 |||||  
 259 roLeuSerAsnHisProLeuAlaGlySerGlyProSerAlaGlyAla 275  
 |||||  
 151 GGCTGCTGCGCGGAGTGCTACCTACCTGGCGCAGGTGGGTGCTGTGACCCG 200  
 |||||  
 276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292  
 |||||  
 201 CACGCCGCTGATGTCTCAGCTGATCGAAGAACCGGTTGCCCGCTCGGTGA 250  
 |||||  
 292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309  
 |||||  
 251 TGCCGGCGGCTGTTGCCGGATCGCTGGTGACGGGTGGCGCGCGCTCCGGTG 300  
 |||||  
 309 etProAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325  
 |||||  
 301 GGTCCGGGACGATGGCCAGGTTCCGAATCCGGCGGCTCCACCAAGCC 350  
 |||||  
 326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342  
 |||||  
 351 GGCTGCTGTCGCCCGCACCGCTCGCGCAGGAGCGGTGAAGAACGACGACG 400  
 |||||  
 342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAspG 359  
 |||||  
 401 AGGACGACTGGGACGAGAGGACTGG 429  
 |||||  
 359 luAspAspTrpAspGluGluAspAspTrp 368

seq\_name: /cgn2\_6/ptodata/2/paa/US086\_COMB.pep:US-08-680-573-109

seq\_documentation\_block:  
; Sequence 109, Application US/08680573  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Campos-Neto, Antonia  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF  
; NUMBER OF SEQUENCES: 128  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/680,573  
; FILING DATE: 12-JUL-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.417C4  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 109:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 368 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-680-573-109

alignment\_scores:  
Quality: 713.00 Length: 143  
Ratio: 5.057 Gaps: 0  
Percent Similarity: 98.601 Percent Identity: 96.503

alignment\_block:  
US-09-462-480-3 x US-08-680-573-109 ..  
Align seg 1/1 to: US-08-680-573-109 from: 1 to: 368

1 CTGCAGCAGGTGACGTGCTGTTTTCAGGCAGGTGGCGGCACCGCGCGG 50  
|||||  
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGly 242  
|||||  
51 CAACCCAGCCGACGAGGAGCGCGCAGATGGCCTGCTCGGCACCAATC 100  
|||||  
242 YAsnProAlaAspGluGluAlaGlnMetGlyLeuLeuGlyThrSerP 259  
|||||  
101 CGCTGTCGACCATCCCTGCTGGTGGTGGATCAGGCCCGCGCGCGG 150  
|||||  
259 roLeuSerAsnHisProLeuAlaGlySerGlyProSerAlaGlyAla 275  
|||||  
151 GGCTGCTGCGCGGAGTGCCTACCTGGCGCGAGTGGTGGTGGTGGCC 200  
|||||  
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlySerLeuThrAr 292  
|||||  
201 CAGCGCGCTGATGTCACGTGATCGAAAGCCGGTTGCCCGCTCGGTGA 250  
|||||  
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309  
|||||

251 TGCCGGCGGCTGTTGCCGATCGTCGGTGACGGTGGCGCGCTCCGGTG 300  
|||||  
309 etProAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325  
|||||  
301 GGTCCGGGAGCATGGCGCAGGTTCGCAATCCGGCGCTCCACCGACC 350  
|||||  
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArg 342  
|||||  
351 GGGTCTGTCGCGCGCGGCACCGCTCGCGCAGGAGCGTGAAGAAGACG 400  
|||||  
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAsp 359  
|||||  
401 AGGACGACTGGGACGACGAGGACGACTGG 429  
|||||  
359 luAspAspTirpAspGluGluAspTirp 368

seq\_name: /cgn2\_6/ptodata/2/paa/US086\_COMB.pep:US-08-680-574-114

seq\_documentation\_block:  
; Sequence 114, Application US/08680574  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Campos-Neto, Antonio  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY  
; NUMBER OF SEQUENCES: 133  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/680,574  
; FILING DATE: 12-JUL-1996  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.411C4  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 114:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 368 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-680-574-114

alignment\_scores:  
Quality: 713.00 Length: 143  
Ratio: 5.057 Gaps: 0  
Percent Similarity: 98.601 Percent Identity: 96.503

alignment\_block:  
US-09-462-480-3 x US-08-680-574-114 ..  
Align seg 1/1 to: US-08-680-574-114 from: 1 to: 368

1 CTGCAGCAGGTGACGTGCTGTTTTCAGGCAGGTGGCGGCACCGCGCGG 50  
|||||  
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGly 242  
|||||

```
51 CAACCCAGCCGACGAGGAGCCGCGCAGATGGGCTGCTCGGCACCCAGTC 100
|||||
242 yAsnProAlaAspGluGluAlaAlaGlnMetGlyLeuGlyThrSerP 259
|||||
101 CGCTGTGCAACCATCCGCTGCTGGTGGATCAGGCCCGCGGGCGG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275
|||||
151 GGCCTGTGCGCGGGAGTCTACCTGCGCAGGTGGTGGTGGTGGTGA 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292
|||||
201 CACGCCGTGATGCTCAGCTGATCGATCGAAGCCGCTGCGCCCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuGluLysProValAlaProSerValM 309
|||||
251 TGCCGGGGCTGTGCGCGATCGCTGCGGTCAGCGGTGGCGCGCTCCGGTG 300
|||||
309 etProAlaAlaAlaGlyGlySerAlaThrGlyAlaAlaProVal 325
|||||
301 GGTCTGTCGCGCGGAGTTCGCAATCCGCGGCTCCACCGCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
|||||
351 GGGTCTGTCGCGCGCACCGCTCGCCAGGAGCGTGAAGAACGACGACG 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAspG 359
|||||
401 AGGACGACTGGGACGAGGACGACTGG 429
|||||
359 luAspAspTrpAspGluGluAspAspTrp 368
|||||

seq_name: /cgn2_6/ptodata/2/paa/US087_COMB.pep:US-08-729-622-109

seq_documentation_block:
; Sequence 109, Application US/08729622
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas H.
; APPLICANT: Twardzik, David R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 132
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; APPLICATION DATA:
; APPLICATION NUMBER: US/08/729,622
; FILING DATE: 11-OCT-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 109:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 368 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-729-622-109

alignment_scores:
Quality: 713.00 Length: 143
Ratio: 5.057 Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503

alignment_block:
US-09-462-480-3 x US-08-729-622-109
Align seg 1/1 to: US-08-729-622-109 from: 1 to: 368

1 CTGACGACGAGTGACGTCGTTGTTTCAGCCAGGTGGGCGGCACCGCGGGCGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGlyG 242
|||||
51 CAACCCAGCCGACGAGGAGCCGCGCAGATGGGCTGCTCGGCACCCAGTC 100
|||||
242 yAsnProAlaAspGluGluAlaAlaGlnMetGlyLeuGlyThrSerP 259
|||||
101 CGCTGTGCAACCATCCGCTGCTGGTGGATCAGGCCCGCGGGCGG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275
|||||
151 GGCCTGTGCGCGGGAGTCTACCTGCGCAGGTGGTGGTGGTGGTGA 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292
|||||
201 CACGCCGTGATGCTCAGCTGATCGATCGAAGCCGCTGCGCCCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuGluLysProValAlaProSerValM 309
|||||
251 TGCCGGGGCTGTGCGCGATCGCTGCGGTCAGCGGTGGCGCGCTCCGGTG 300
|||||
309 etProAlaAlaAlaGlyGlySerAlaThrGlyAlaAlaProVal 325
|||||
301 GGTCTGTCGCGCGGAGTTCGCAATCCGCGGCTCCACCGCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
|||||
351 GGGTCTGTCGCGCGCACCGCTCGCCAGGAGCGTGAAGAACGACGACG 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAspG 359
|||||
401 AGGACGACTGGGACGAGGACGACTGG 429
|||||
359 luAspAspTrpAspGluGluAspAspTrp 368
|||||

seq_name: /cgn2_6/ptodata/2/paa/US087_COMB.pep:US-08-729-622-109

seq_documentation_block:
; Sequence 114, Application US/08730510
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 137
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
```

```
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/730,510
; FILING DATE: 27-AUG-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 114:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 368 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-730-510-114

alignment_scores:
  Quality: 713.00      Length: 143
  Ratio: 5.057        Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503

alignment_block:
US-09-462-480-3 x US-08-730-510-114
..
Align seg 1/1 to: US-08-730-510-114 from: 1 to: 368

1 CTGAGCAGGTGACGTCGTTGTTACGACGAGTGGCGGCACCGCGCGG 50
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGlyG 242
51 CAACCCAGCGCAGGAGGAGCGGCAGATGGCTGCTCGGCACCGAGTC 100
242 yAsnProAlaAspGluGluAlaGlnMetGlyLeuLeuGlyThrSerP 259
101 CGCTGTGCAACATCCGCTGGCTGGTGGTGGTGGTGGTGGTGGTGG 150
259 roLeuSerAsnHisProLeuAlaGlySerGlyProSerAlaGlyAla 275
151 GGCCTGTGCGCGGAGTCGCTACCTGGCGCAGGTGGTGGTGGTGGT 200
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThr 292
201 CACGCGCTGATGCTCAGCTGATCGAAGCGGTGCTCCCGCTCGGTGA 250
292 gThrProLeuMetSerGlnLeuLeuGluProValAlaProSerValM 309
251 TCGCGCGCGCTGTTGCCGATGCTCGGTGACGGTGGCGCGCTCGGTG 300
309 etProAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
301 GGTCTGGAGCGATGGCGCAGGGTTCGCAATCCCGCGGCTCCACAGCCC 350
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArg 342
351 GGGTCTGTGTCGCGCGCAGCGCTCGCGCAGGAGCGGTGCAAGACGAC 400
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluAspAspG 359
401 AGGACGACTGGGACGAGGAGGACGACTGG 429
359 luAspAspTrpAspGluAspAspTrp 368
```

```
seq_name: /cgn2_6/ptodata/2/paa/US089_COMB.pep.US-08-942-341-109

seq_documentation_block:
; Sequence 109, Application US/08942341
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 209
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/942,341
; FILING DATE: 01-OCT-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C7
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 109:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 368 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-942-341-109

alignment_scores:
  Quality: 713.00      Length: 143
  Ratio: 5.057        Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503

alignment_block:
US-09-462-480-3 x US-08-942-341-109
..
Align seg 1/1 to: US-08-942-341-109 from: 1 to: 368

1 CTGACGAGGTGACGTCGTTGTTACGACGAGTGGCGGCACCGCGCGG 50
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGlyG 242
51 CAACCCAGCGCAGGAGGAGCGGCAGATGGCTGCTCGGCACCGAGTC 100
242 yAsnProAlaAspGluGluAlaGlnMetGlyLeuLeuGlyThrSerP 259
101 CGCTGTGCAACATCCGCTGGCTGGTGGTGGTGGTGGTGGTGGTGG 150
259 roLeuSerAsnHisProLeuAlaGlySerGlyProSerAlaGlyAla 275
151 GGCCTGTGCGCGGAGTCGCTACCTGGCGCAGGTGGTGGTGGTGGT 200
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThr 292
201 CACGCGCTGATGCTCAGCTGATCGAAGCGGTGCTCCCGCTCGGTGA 250
292 gThrProLeuMetSerGlnLeuLeuGluProValAlaProSerValM 309
251 TCGCGCGCGCTGTTGCCGATGCTCGGTGACGGTGGCGCGCTCGGTG 300
309 etProAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
301 GGTCTGGAGCGATGGCGCAGGGTTCGCAATCCCGCGGCTCCACAGCCC 350
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArg 342
351 GGGTCTGTGTCGCGCGCAGCGCTCGCGCAGGAGCGGTGCAAGACGAC 400
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluAspAspG 359
401 AGGACGACTGGGACGAGGAGGACGACTGG 429
359 luAspAspTrpAspGluAspAspTrp 368
```

```
201 CACGCCGCTGATGCTCAGCTGATCGAAGAACCGGTTGCCCGCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309
|||||
251 TCCCGCGCGCTGTCGCCGATCGCGGTGACGGGTGGCGCGCTCCCGGTG 300
|||||
309 etProAlaAlaAlaGlySerAlaThrGlyGlyAlaAlaProVal 325
|||||
301 GGTCTGCTGCGGAGCGGCGGCTCGCAATCCGCGGCTCCACGAGCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
|||||
351 GGGTCTGCTGCGCGCGGCGGCTCGCGAGGAGCGGTGAAGAACGACG 400
|||||
342 oglyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspSpG 359
|||||
401 AGGAGCTGGGACGAGACGACGCTGG 429
|||||
359 luAspAspTrpAspGluGluAspTrp 368
|||||
```

seq\_name: /cgn2\_6/ptodata/2/paa/US089\_COMB.pep:US-08-942-578-114

```
seq_documentation_block:
; Sequence 114, Application US/08942578
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skelky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 214
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/942,578
; FILING DATE: 01-OCT-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C7
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 114:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 368 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-942-578-114
```

alignment\_scores:      Quality: 713.00      Length: 143  
                         Ratio: 5.057            Gaps: 0  
Percent Similarity: 98.601      Percent Identity: 96.503

```
alignment_block:
US-09-462-480-3 x US-08-942-578-114
Align seg 1/1 to: US-08-942-578-114 from: 1 to: 368
1 CTGACAGAGTGACGTGCTGTTTCAGCACAGGTGGCGGCGCACCGCGCGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGlyG 242
|||||
51 CAACCCAGCCGACGAGGAGCGCGCAGATGGGCTGCTCGGCACACAGTC 100
|||||
242 yaSProAlaAspGluGluAlaGlnMetGlyLeuGlyLeuGlyThrSerP 259
|||||
101 CGCTGTGCAACCATCCGCTGCTGGTGGATCAGGCCCGCCAGCGCGGCG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275
|||||
151 GGCTGTGTCGCGCGGAGTCCCTACCTGCGCGCAGGTGGTGGTCTGTCG 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThr 292
|||||
201 CACGCCGCTGATGCTCAGCTGATCGAAGAACCGGTTGCCCGCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309
|||||
251 TCCCGCGCGCTGTTCCCGGATCGCTGCGGTGACGGGTGGCGCGCTCCCG 300
|||||
309 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
|||||
301 GGTCCGGAGCGATGGCGGCGGAGGTTGGCAATCCGCGGCTCCACGAGCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
|||||
351 GGGTCTGCTGCGCGCGGCGGCGGCTCGCGAGGAGCGGTGAAGAACGACG 400
|||||
342 oglyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspSpG 359
|||||
401 AGGAGCTGGGACGAGACGACGCTGG 429
|||||
359 luAspAspTrpAspGluGluAspTrp 368
|||||
```

seq\_name: /cgn2\_6/ptodata/2/paa/US090\_COMB.pep:US-09-024-753-109

```
seq_documentation_block:
; Sequence 109, Application US/09024753
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skelky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; NUMBER OF SEQUENCES: 236
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/09/024,753
; FILING DATE: 18-FEB-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
```

TUBERCULOSIS

```
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C8
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 109:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 368 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-09-024-753-109

alignment_scores:
    Quality: 713.00      Length: 143
    Ratio: 5.057        Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503

alignment_block:
US-09-462-480-3 x US-09-024-753-109 ..

Align seg 1/1 to: US-09-024-753-109 from: 1 to: 368

1 CTGCAGCAGGTGACGTGCTGTTGTTCAGCCAGGTGGCGGCACCGCGCGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGlyG 242

51 CAACCCAGCGGACGAGGAGCGCCGACATGGGCTGCTCGGCACCACTC 100
|||||
242 YasnProAlaAspGluGluAlaAlaGlnMetGlyLeuLeuGlyThrSerP 259

101 CGCTGTGGAACCATCCGCTGGCTGGTGGATCAGGCCCGCCAGCGCGCG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275

151 GGCCTGTCTGCGCGGAGTCGCTACCTGGCGCAGGTGGGTGCTTGACCCG 200
|||||
276 GlyLeuLeuArgAlaGlySerAlaThrGlyGlyAlaAlaProVal 292

201 CAGCGCGCTGATGTCAGCTGATCGAATCCGCGCGCTCCACCGCCG 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342

351 GGGTGTGTCGCGCGGACCGCTCGCGCAGGAGCGGTGAAGAAGCAGCG 400
|||||
342 oglyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspG 359

401 AGGACGACTGGGACGAGGAGGAGCTGG 429
|||||
359 luAspAspIrrpAspGluGluAspIrrp 368

seq_name: /cgn2_6/ptodata/2/paa/US090_COMB.pep:US-09-025-197-114

seq_documentation_block:
; Sequence 114, Application: US/09025197
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
```

351 GGGTCTGGTCGGCGGACCGCTCGCGAGGAGCGTGAAGAACGACGACG 400  
|||||  
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspSpG 359  
401 AGGACGACTGGGACGAAGACGACTGG 429  
359 luAspSpTrpAspGluGluAspSpTrp 368

seq\_name: /cgn2\_6/ptodata/2/paa/US090\_COMB.pep:US-09-072-596-109

seq\_documentation\_block:  
; Sequence 109, Application US/09072596  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Campos-Neto, Antonia  
; APPLICANT: Houghton, Raymond  
; APPLICANT: Veddzik, Thomas S.  
; APPLICANT: Twardzik, Daniel R.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Hendrickson, Ronald C.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF  
; TUBERCULOSIS  
; NUMBER OF SEQUENCES: 350  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/072,596  
; FILING DATE: 05-MAY-1998  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.417C9  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 109:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 368 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-09-072-596-109

alignment\_scores:  
Quality: 713.00 Length: 143  
Ratio: 5.057 Gaps: 0  
Percent Similarity: 98.601 Percent Identity: 96.503

alignment\_block:  
US-09-462-480-3 x US-09-072-596-109 ..

Align seg 1/1 to: US-09-072-596-109 from: 1 to: 368

1 CTGACGAGGTACGTCGTTGTTCAGCGAGTGGCGGACCGCGCGG 50  
|||||  
226 LeuGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGly 242  
51 CAACCCAGCCGACGAGGAGCGGCGGACGATGGCGCTGCTGGCACCACTC 100  
|||||

242 yAnProAlaAspGluGluAlaAlaGlnMetGlyLeuLeuGlyThrSerP 259  
101 CGCTGTGCAACCATCCGCTGGTGGATCAGGCCCGCCAGCGGGCGGCG 150  
|||||  
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275  
151 GGCTGTGTCGCGCGGAGTCTACCTGCGCGAGGTGGTGGTGGTGGTGG 200  
|||||  
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292  
201 CAGCCCGCTGATGTCTCAGCTGATCGAAAGCCGGTTGCCCTCGGTGA 250  
|||||  
292 gThrProLeuMetSerGlnLeuLeGluLysProValAlaProSerValM 309  
251 TGC CGCGGCGCTGTTCCGCGATCGTCGTCGACGGTGGCGCGCTCCGGTG 300  
|||||  
309 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325  
301 GTTCCGGGAGCGATGGGCCAGGGTTCCCAATCCGGCGGCTCCACGACCC 350  
|||||  
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342  
351 GGCTCTGTCGCGCGGCGCGCTCGCGAGGCGTGAAGAACGACGACG 400  
|||||  
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspSpG 359  
401 AGGACGACTGGGACGAAGACGACTGG 429  
359 luAspSpTrpAspGluGluAspSpTrp 368

seq\_name: /cgn2\_6/ptodata/2/paa/US090\_COMB.pep:US-09-072-967-114

seq\_documentation\_block:  
; Sequence 114, Application US/09072967  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Campos-Neto, Antonio  
; APPLICANT: Houghton, Raymond  
; APPLICANT: Veddzik, Thomas S.  
; APPLICANT: Twardzik, Daniel R.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Hendrickson, Ronald C.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY  
; AND DIAGNOSIS OF TUBERCULOSIS  
; NUMBER OF SEQUENCES: 355  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/072,967  
; FILING DATE: 05-MAY-1998  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.411C9  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 114:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 368 amino acids

```
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-072-967-114

alignment_scores:
  Quality: 713.00      Length: 143
  Ratio: 5.057        Gaps: 0
  Percent Similarity: 98.601  Percent Identity: 96.503

alignment_block:
  US-09-462-480-3 x US-09-072-967-114

Align seg 1/1 to: US-09-072-967-114 from: 1 to: 368

1 CTGCAGCAGGTGACGTCGCTGTTTCAGCCAGGTGGCGGCACCGCGCGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGlyG 242
51 CAACCCAGCCGACGAGGAAGCGCGCAGATGGCTGCTCGGCACCGATC 100
|||||
242 YAsnProAlaAspGluGluAlaGlnMetGlyLeuLeuGlyThrSerP 259
101 CGCTGTCGAACCATCCGCTGGCTGGATCATCGGCCCGCGCGCGCG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275
151 GGCCTGCTGCGCGCGGAGTCGCTACCTGGCGCAGGTGGTGTGACCCG 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292
201 CACGCCGCTGATGCTCAGCTGATCGATCGAAGACCGCGTTGCCCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuLeuGluProValAlaProSerValm 309
251 TGGCGCGCGCTGTGCGGATCGTCGGTGACGGTGCGCGCGCTCGCGGTG 300
|||||
309 etProAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
301 GGTCCGGGAGCGATGGCGGAGGTTCGCAATCCGCGCGCTCCACGACCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
351 GGGTGTGTCGCGCGCGCACGCTCGCGCAGGAGCGTGTAAGAACGACGACG 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAspG 359
401 AGGAGGACTGGGACGAGAGGACGACTGG 429
359 luAspAspTrpAspGluGluAspTrp 368

seq_name: /cgn2_6/ptodata/2/paa/US097_COMB.pap.us-09-724-685-114

seq_documentation_block:
; Sequence 114, Application US/09724685
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; Skeiky, Yasir
; Dillon, Davin C.
; Campos-Neto, Antonio
; TITLE OF INVENTION: Compounds and Methods for
; Immunotherapy and Diagnosis of Tuberculosis
; NUMBER OF SEQUENCES: 155
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/724,685
; FILING DATE: 28-Nov-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/523,436
; FILING DATE: 01-SEP-1995
; APPLICATION NUMBER: US 08/533,634
; FILING DATE: 22-SEP-1995
; APPLICATION NUMBER: US 08/620,874
; FILING DATE: 22-MAR-1996
; APPLICATION NUMBER: US 08/659,683
; FILING DATE: 05-JUN-1996
; APPLICATION NUMBER: US 08/680,574
; FILING DATE: 12-JUL-1996
; APPLICATION NUMBER: WO PCT/US96/14674
; FILING DATE: 30-AUG-1996
; APPLICATION NUMBER: US 08/730,511
; FILING DATE: 11-OCT-1996
; APPLICATION NUMBER: US 08/818,112
; FILING DATE: 13-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Bastian, Kevin L.
; REGISTRATION NUMBER: 34,774
; REFERENCE/DOCKET NUMBER: 014058-008561US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 114:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 368 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 114:
US-09-724-685-114

alignment_scores:
  Quality: 713.00      Length: 143
  Ratio: 5.057        Gaps: 0
  Percent Similarity: 98.601  Percent Identity: 96.503

alignment_block:
  US-09-462-480-3 x US-09-724-685-114

Align seg 1/1 to: US-09-724-685-114 from: 1 to: 368

1 CTGCAGCAGGTGACGTCGCTGTTTCAGCCAGGTGGCGGCACCGCGCGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGlyG 242
51 CAACCCAGCCGACGAGGAAGCGCGCAGATGGCTGCTCGGCACCGATC 100
|||||
242 YAsnProAlaAspGluGluAlaGlnMetGlyLeuLeuGlyThrSerP 259
101 CGCTGTCGAACCATCCGCTGGCTGGATCATCGGCCCGCGCGCGCG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275
151 GGCCTGCTGCGCGCGGAGTCGCTACCTGGCGCAGGTGGTGTGACCCG 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292
201 CACGCCGCTGATGCTCAGCTGATCGATCGAAGACCGCGTTGCCCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuLeuGluProValAlaProSerValm 309
251 TGGCGCGCGCTGTGCGGATCGTCGGTGACGGTGCGCGCGCTCGCGGTG 300
|||||
309 etProAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
301 GGTCCGGGAGCGATGGCGGAGGTTCGCAATCCGCGCGCTCCACGACCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
351 GGGTGTGTCGCGCGCGCACGCTCGCGCAGGAGCGTGTAAGAACGACGACG 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAspG 359
401 AGGAGGACTGGGACGAGAGGACGACTGG 429
359 luAspAspTrpAspGluGluAspTrp 368
```





**THIS PAGE BLANK (USPTO)**





```
114 ....AsnProProSerProLeuProAlaProAlaProLeuArgA 129
||||| ||| ||||| ||| |||||
282 CGTCACGACGACCGCGCAACAGCCGCGCATCACCGAGGGGCAACCG 233
||||| ||||| ||||| ||||| ||||| |||||
129 laSerProLeu.....SerProThrThrSerProSerLeuPro 142
||||| ||||| ||||| ||||| ||||| |||||
232 GCTTTTCGATCAGTACGATCAGCGCGTGGGTCAACGACCCACCT 183
||||| ||||| ||||| ||||| ||||| |||||
143 LeuProProSerSerAlaAsnAlaPheAlaLeuThrThr..Pro 158
||||| ||||| ||||| ||||| ||||| |||||
182 GCGCCAGTAGCGACTCGCGCGCAGCGCC.....GCGCCCGC 142
||| ||| ||||| ||||| |||||
159 ProProLysThrThrProSerArgSerSerProProArgArgProPr 175
||||| ||||| ||||| ||||| ||||| |||||
141 GCTGGGCGCTGATCCACCGACCGCGGATGTTTCGACAGCGGAGTGTGC 92
||||| ||||| ||||| ||||| ||||| |||||
175 oSerGlyProLysLeuProAlaProThrSerAlaArgSerGlyAlaSerP 192
|| ||||| ||||| ||||| ||||| |||||
91 CGAGCAGGCC...ATCTGCGCGGCTTCCTCGTGGGTGGGTGC..... 50
|| ||||| ||||| ||||| ||||| |||||
192 roProArgProAsnSerCys.ArgCysArgArgArgileGlnCysPheTh 208
|| ||||| ||||| ||||| ||||| |||||
49 .....GCGCCCGCGTGGCGCCACCT..... 29
||| ||||| ||||| ||||| |||||
208 rIleThrIleAsnThrThrSerSerAsnSerAsnArgProProTrpGluA 225
||| ||||| ||||| ||||| |||||
28 .....GGCTGACACACGACGTCACCTGCT 5
||||| ||||| ||||| ||||| |||||
225 rgProLeuProGlnGlySerGlyIleThrSerProAla 237
```

seq\_name: /cgn2\_6/ptodata/2/paa/US60\_NEW\_COMB.pep:US-60-382-898-278

```
seq_documentation_block:
; Sequence 278, Application US/60382898
; GENERAL INFORMATION:
; APPLICANT: Hudson, Keith
; APPLICANT: et al.
; TITLE OF INVENTION: Plant Receptors and Ligands
; FILE REFERENCE: 1066P
; CURRENT APPLICATION NUMBER: US/60/382,898
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 1344
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 278
; LENGTH: 708
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-60-382-898-278
```

alignment\_scores:  
Quality: 125.50 Length: 175  
Ratio: 1.793 Gaps: 7  
Percent Similarity: 40.000 Percent Identity: 29.143

alignment\_block:

US-09-462-480-3/rev x US-60-382-898-278 ..

Align seg 1/1 to: US-60-382-898-278 from: 1 to: 708

```
421 CCTTTGTCCTCCAGTCGTCCTCGTCTTCTTCACGCTCTCGCGGAGC 372
||||| ||||| ||||| ||||| ||||| |||||
38 ProLeuProProSerAlaPro.....ProProAsnArgAl 49
||||| ||||| ||||| ||||| ||||| |||||
371 GGTGCGCGCGGACGACCGCGGCTGGTGAGCCGCGCGGATTCGGAAC. 323
||| ||| ||||| ||||| ||||| |||||
49 aProProProProProValThrThrSerProProProValAlaAsnG 66
||||| ||||| ||||| ||||| ||||| |||||
322 .....CCTGGCCCATCGCTCCGAGCCACCGGAGCGGCCAC..... 284
||||| ||||| ||||| ||||| ||||| |||||
66 lAlaProProProProLeuProLysProProGluSerSerProPro 82
```

```
283 CCCTACCGACGATCCGCAACAGCCCGCGCA...TCACCGAGGGGCA 237
||||| ||||| ||||| ||||| ||||| |||||
83 ProGlnProValIleProSerProProSerThrSerProProGln 99
||||| ||||| ||||| ||||| ||||| |||||
236 ACCGGCTTTTCGATCAGTACGATCAGACATCAGCGCGGTG..... 201
||||| ||||| ||||| ||||| ||||| |||||
99 nProValIleProSerProProProSerAlaSerProProAlaLeuV 116
||||| ||||| ||||| ||||| ||||| |||||
201 ..... 201
116 alProProLeuProSerSerProProProProAlaSerValProPro 132
||||| ||||| ||||| ||||| ||||| |||||
201 ..... 201
133 ArgProSerProSerProProIleLeuValArgSerProProSerVa 149
||||| ||||| ||||| ||||| ||||| |||||
200 ....CGGGTCAACAGACCCACCTGGCGCCAGGTAGCGATCCGCGCGCAGCA 155
||||| ||||| ||||| ||||| ||||| |||||
149 lArgProIleGlnSerProProProProSerAspArgProThrGlnS 166
||||| ||||| ||||| ||||| ||||| |||||
154 GCGCCGCGCGCGCTGGGCGCTGATCCACCGACCGCGGATGTTTCGAC 105
||||| ||||| ||||| ||||| ||||| |||||
166 erProProProProSerProProSerGluArgProThrGln 182
||||| ||||| ||||| ||||| ||||| |||||
104 AGCGGACTGTGTCGAGCAGGCCCATCTGCGCGCTTCCTCGTCGGCTGG 55
||| ||||| ||||| ||||| ||||| |||||
183 Ser.....ProProSerProProSerGluArgProThrGlnSerPro.. 196
||||| ||||| ||||| ||||| ||||| |||||
54 GTTGCCCGCGCGGTGGCGCCACCC 30
||||| ||||| ||||| ||||| ||||| |||||
197 ....ProProSerProProSer 203
```

seq\_name: /cgn2\_6/ptodata/2/paa/US10\_NEW\_COMB.pep:US-10-155-881-25973

```
seq_documentation_block:
; Sequence 25973, Application US/10155881
; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu Jingdong
; APPLICANT: Lutfiyya, Linda L.
; APPLICANT: McIninch, James
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH TRANSCRIPTION IN PLANTS
; FILE REFERENCE: 38-21(15300)J
; CURRENT APPLICATION NUMBER: US/10/155,881
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 37595
; SEQ ID NO 25973
; LENGTH: 318
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-155-881-25973
```

alignment\_scores:  
Quality: 125.00 Length: 187  
Ratio: 1.562 Gaps: 11  
Percent Similarity: 42.781 Percent Identity: 29.947

alignment\_block:

US-09-462-480-3 x US-10-155-881-25973 ..

Align seg 1/1 to: US-10-155-881-25973 from: 1 to: 318

```
12 GAGCTCGTCTTTCAGCCAGGTGGCGCCAGCGCGCGGCAACCCAGCCG 61
||||| ||||| ||||| ||||| ||||| |||||
120 AspAlaPheProArgAlaGlyAspArgGlnArgArgHis.....Ar 134
||||| ||||| ||||| ||||| ||||| |||||
62 ACAGGAGAGCGCGCAGATGGCGCTGCTCGCACCAGTCGCTGCTCGAAC 111
||||| ||||| ||||| ||||| ||||| |||||
134 gArgGly.....HisArgGlyLeuLeuProp 143
```



```

; APPLICANT: AMSHEY, Stefan R.
; APPLICANT: DAUGHERTY, Sean C.
; APPLICANT: DAM, Tam C.
; APPLICANT: LIU, Tommy F.
; APPLICANT: NGUYEN, Duy-Viet An
; APPLICANT: KLEPFELD, Yael
; APPLICANT: GERSTIN JR., Edward H.
; APPLICANT: PERALTA, Careyana H.
; APPLICANT: DAVID, Marie H.
; APPLICANT: LEWIS, Samantha A.
; APPLICANT: CHEN, Alice J.
; APPLICANT: PANZER, Scott R.
; APPLICANT: HARRIS, Bernard
; APPLICANT: FLORES, Vincent
; APPLICANT: MARWAHA, Rakesh
; APPLICANT: LO, Audrey
; APPLICANT: LAN, Ruth Y.
; APPLICANT: URASHKA, Michael
; TITLE OF INVENTION: SECRETORY MOLECULES
; FILE REFERENCE: PT-1232 PCT
; CURRENT APPLICATION NUMBER: PCT/US02/09921
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: 60/280,067; 60/280,068; 60/291,280; 60/291,849;
; 60/291,829; 60/299,428; 60/300,001; 60/299,776
; PRIOR FILING DATE: 2001-03-29; 2001-03-29; 2001-05-16; 2001-05-17;
; 2001-05-17; 2001-06-19; 2001-06-20; 2001-06-20
; NUMBER OF SEQ ID NOS: 1146
; SOFTWARE: PERL Program
; SEQ ID NO 1022
; LENGTH: 380
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: LG:007916.8.orf2:2001JUN22
PCT-US02-09921-1022
```

```

alignment_scores:
  Quality: 121.50      Length: 167
  Ratio: 1.293        Gaps: 10
  Percent Similarity: 56.287      Percent Identity: 34.132
```

```

alignment_block:
US-09-462-480-3/rev x PCT-US02-09921-1022 ..
```

Align seg 1/1 to: PCT-US02-09921-1022 from: 1 to: 380

```

479 AGTCTCCGCGCGGGTGGCGGGAAGCTCTGTTGTCATTACGGGAGCTCA 430
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
137 SerLeuTyrAlaAspSerAspGlyGluAsnAspSerThrSerSerSergI 153
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
429 CCAGTCGCTCTTCGTCGCCAGTCGCTCGTCGCTCTTCTTCACGCTCCT 380
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
153 uSerSerSerSerSerSerSerSerSerSerSerSerSerSerSers 170
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
379 CGCGGACGGTGGCGGCGGACACCGCGGCTGTGGAGCCCGCCGGAT 330
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
170 erSerSerSerSerSerSerSerSerSerSerSerSerSerSerSers 186
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
329 TGGCAACCTGGCCCATCGTCCCG .....GACCCACCGAGCGGCGCC 286
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
187 GluGlu.ArgProAlaAlaLeuProSerAlaSerProProArgGluV 203
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
285 ACCCGTCACCGACG .....ATCCGCGCAACAGCGC 257
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
203 alProValProThrProAlaProValGluValProValProGluArgVal 219
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
256 CCGGCATACCGAGGGGCAACCGGCTTTTGATCAGTCAGACATCAGC 207
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
220 AlaGlySerProValThrProProAlaArgThrGlyGlyValSerSerLy 236
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
206 GCGGTGC .....GGGTCAACGACCCACCTCGCGCCAGGTACCGACTC 166
```

```

||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
236 sAlaCysArgProHisGlyValThrProGlnCysAlaSerAlaSerP 253
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
165 CGCGCGCGACGA .....GGCCGCGCGCGCTGGGCTGATCCACGAG 122
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
253 roArgThrThrCysTrpAlaProArgPro .....LeuProHisAla 266
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
121 CCAGCGGATGTTGCAGACGCGGACTGTGCGGACGAGCCCATCTGCGCG 72
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
267 ProMet .....SerValPro..LeuLeuProSer...ProSerCysPro 279
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
71 GCTTCCTCGTCGCGCTGGG .....TTGCCGCGCGCGTGGCGCCACCTGG 27
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
280 HisProArgAsnAlaGlyLysLeuSerProSerLeuProSerArgTrp 295
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
seq_name: /cgn_6/ptodata/2/paa/US60_NEW_COMB.pep:US-60-360-039-3962

seq_documentation_block:
; Sequence 3962, Application US/60360039
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Chen, Xianfeng
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)A
; CURRENT APPLICATION NUMBER: US/60/360,039
; CURRENT FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 3962
; LENGTH: 639
; TYPE: PRT
; ORGANISM: Neurospora crassa
US-60-360-039-3962
```

```

alignment_scores:
  Quality: 121.50      Length: 189
  Ratio: 1.538        Gaps: 12
  Percent Similarity: 41.799      Percent Identity: 31.217
```

```

alignment_block:
US-09-462-480-3/rev x US-60-360-039-3962 ..
```

Align seg 1/1 to: US-60-360-039-3962 from: 1 to: 639

```

473 CGCGCGCGGGTGGCGGGAAGTCTGTTGTCATTACGGGAGCTCACCAGTC 424
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
319 ProAlaArgArgSerGlyLys .....LeuAspThrGluAsnHisGln.G 333
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
423 GTCCTCTTCGTCGCCAGTCGT ..... 404
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
333 luProAlaProProProArgPheAlaValProProProIleAlaAspAla 349
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
403 .....CCTCGTGTCTTCTTCACGTCCTCTGC 378
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
350 GlyLysPheAlaHisSerAspProProArgHisThrProSerAlaProGl 366
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
377 GCGAGCGGTGCCGCGCGGACCA .....GACC 352
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
366 yProProProProArgProProLysThrProLeuGluAspGlnAspP 383
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
351 CGGCTGTGGAGCGCGGATTCGGAACCTGCGCCA ..... 314
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
383 ro .....SerGlnArgPheSerValProProPheThrGlyCln 396
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
313 ...TCGCTCCCGGACCCACCGAGCGGCCAC .....CGGTCCAC 276
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
397 ArgSerValProProProProSerArgSerValProProProPr 413
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
275 GACGATCCGGCAA .....CAGCGCGCGGCATCACCGAGGGGCAAC 235
```

413 oProArgAsnSerAlaAlaGlnProProLeuProProLys..AlaPr 429  
234 CGGCTTTTCAGTCAGTCAGCATCAGCGGCTGCGGTCAACGACCAC 185  
429 oGly.....P 431  
184 CTGGCCGAGTAGGACTCCGCGCGCAGCAGCGCCGCGCGCGCTG... 138  
431 roAlaProProLeuProProAlaSerArgProProMetLeuPro 447  
137 .....GGGCTGATCCACGACGCGGATGTTGACACGCGACTGGT 94  
448 ThrArgSerProAlaProProGlnAlaProProLeuProThrSerAsnAl 464  
93 GCGGAGCAGGCCCATCTGCGCGGCTTCTCTGCTGCGGTGGGTGCGCGC 44  
464 aprProProProLeuProAla....ThrGlnAlaProProProp 480  
43 CGTGCGCGCCGCC 30  
480 roLeuProAlaThr 484

seq\_name: /cgn2\_6/ptodata/2/paa/US09\_NEW\_COMB.pep.us-09-784-984A-8

seq\_documentation\_block:  
; Sequence 8, Application US/09784984A  
; GENERAL INFORMATION:  
; APPLICANT: AUDONNET et al.  
; TITLE OF INVENTION: POLYNUCLEOTIDE VACCINE FORMULA AGAINST PORCINE  
; TITLE OF INVENTION: REPRODUCTIVE AND RESPIRATORY PATHOLOGIES  
; FILE REFERENCE: 454313-2230  
; CURRENT APPLICATION NUMBER: US/09/784,984A  
; CURRENT FILING DATE: 2001-02-16  
; PRIOR APPLICATION NUMBER: US/09/232,468A  
; PRIOR FILING DATE: 1999-01-05  
; NUMBER OF SEQ ID NOS: 54  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 8  
; LENGTH: 404  
; TYPE: PRT  
; ORGANISM: Aujeszky's Disease Virus (NIA3 Strain)  
US-09-784-984A-8

alignment\_scores:  
Quality: 121.00 Length: 166  
Ratio: 1.424 Gaps: 7  
Percent Similarity: 51.205 Percent Identity: 30.723

alignment\_block:  
US-09-462-480-3/rev x US-09-784-984A-8 ..

Align seg 1/1 to: US-09-784-984A-8 from: 1 to: 404

472 CGGCGCGGTCGGCGGGAAGTCTGTTGTCATTACGGGAG...CTCACCAG 426  
193 ArgProArgGlyProAlaProHisValGlnValArgArgValLeuGluAr 209  
425 TCGTCTCTCTCCCGAGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 376  
209 gArgGlnLeuGlnAlaGlyArgGlyArgGlyArgGlyArgGlyArgGly 224  
375 GAGGGTTCGCGCGCAGCAGCCGCGGCTGTTGAGCGCGCGGATGCG 326  
224 laValLeuProAlaProAlaProGlyGlyGlyGlyGlyGlyGlyGlyGly 240  
325 AACCTGCGCCATCGCTCCCGGAGCCACCGAGCGGCGCCACCGCTACC 276  
241 GlnGluArgProArgProAlaProAlaGlyProArgArgHis..... 254  
275 GAGCATCGGCAACAGCGCGGCA.....TCAC 247

255 AlaValArgHisArgProArgAlaAlaLeuGlyGlyLeuAlaGluAlaP 271  
246 CGAGGGGCAACCGGCTTTTCATCAGCTGAGCATCATCAGCGCGTGGCG 197  
271 roAlaProAlaProAlaProAlaGluAlaArgAlaArgProGly 287  
196 TCACAGCCACCTGCGCGCAGGTAGCGACTCCGCGCGCAGCAGCGCCGCG 147  
288 AspAlaArgAlaProArgProProAlaArgAlaGlyAspAlaGlyProAr 304  
146 CCGCGCTGGGCGCTGATCCACGCGGATGGTTCGACAGCGGACT 97  
304 gArgArg.....GlyProProHisAlaAlaThrProGluA 316  
96 GGTGCGCGCAGCGC.....CCATCTGCGCGGCTTCTCTGTCGG 59  
316 laArgAspAlaAlaProProLeuArgProAlaGlyArgAlaGlnArg 332  
58 CTGGTTCGCGCGCGGTCGCCACCTCGCTGGTGAACAACGAGCT 13  
333 ValAlaAlaAlaArgGlyAlaValProAlaAlaAspProArgArg 347

seq\_name: /cgn2\_6/ptodata/2/paa/US10\_NEW\_COMB.pep.us-10-155-881-11183

seq\_documentation\_block:  
; Sequence 11183, Application US/10155881  
; GENERAL INFORMATION:  
; APPLICANT: Dotson, Stanton B.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Lutfiyya, Linda L.  
; APPLICANT: McIninch, James  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH  
; TITLE OF INVENTION: TRANSCRIPTION IN PLANTS  
; FILE REFERENCE: 38-21(15300)J  
; CURRENT APPLICATION NUMBER: US/10/155,881  
; CURRENT FILING DATE: 2002-05-22  
; NUMBER OF SEQ ID NOS: 37595  
; SEQ ID NO 11183  
; LENGTH: 668  
; TYPE: PRT  
; ORGANISM: Zea mays  
US-10-155-881-11183

alignment\_scores:  
Quality: 121.00 Length: 178  
Ratio: 1.344 Gaps: 8  
Percent Similarity: 50.562 Percent Identity: 29.775

alignment\_block:  
US-09-462-480-3/rev x US-10-155-881-11183 ..

Align seg 1/1 to: US-10-155-881-11183 from: 1 to: 668

481 CAATCTTCGCGCGCGGTCGGGGAAGTCTGTTGTCATTACGGGAGCT 432  
171 GluIleIleHisProCysAsnProGly...LeuAlaSerLeuLeuGluLe 186  
431 CACGAGTCGCTCTT.....CGTCCCA 409  
186 uArgLeuArgSerLeuLeuAlaAlaAspProAlaProLeuProProp 203  
408 GTCCTCTCTGTCGT.....CTTCTTCACGCTCTCTGCGCGAGCGGTGCG 365  
203 roGlnProGlnHisAlaLeuLeuHisGlyAlaProAlaAlaPro 219  
364 GCG.....CGACGACAGCCGCGGTGTGTGAGCCCGCGGATTCGAA 324  
220 AlaGlyLeuThrLeuProPro..... 227  
323 CCCTGGCCCATCGTCCGCGACCCACCGAGCGCGCCACCGCTCACCGA 274



```
228 .....ProProLeuProAspLysArgArgHisGluHisProProProc 242
273 CGATCCGGCAACAG.....CCGGCGGATCACCAGGGGGGCAACCG 233
      :|||||
242 ysgInGlnGlnGlnGlnGlnProHisProAlaProGlnSerProLys 258
      :|||
232 GCTTTTCGATCGTGTAGACATCAGCGCGGTGCGGGTCAACAGCCACCT 183
      :|||
259 AlaProThrAlaGluGluThrAlaAlaAlaAlaAla..... 270
182 GCGCCAGGTAGCACTCCGGCGGCGAGCGCGCGCGCGCGTGGGGCC 133
      :|||
271 .....AlaAlaGlnAlaAlaAlaAlaAlaAlaAla 280
132 TGATCCACCAGCAGCGGATGTTTCGACAGCGGACTGGTCCGAGCAGGC 83
      :|||
281 LysGluArgLysGluGluGlnArgArgLysGlnArgAspGluGluGlyLe 297
82 CCATCTCGCGGGTCTCTCGTGGCTGGTTCGCGCGCGCGGTGCCGCC 33
      :|||
297 uHisLeuThrLeuLeuLeuGlnCysAlaGluAlaValAlaAlaAspA 314
32 ACCTGGCTGAACACGACGTCACCTGCTGCAG 1
314 snLeuAspAspAlaHisGlnThrLeuLeuGlu 324
```

seq\_name: /cgn2\_6/ptodata/2/paa/US10\_NEW\_COMB.pep:US-10-155-881-10047

seq\_documentation\_block:

; Sequence 10047, Application US/10155881

; GENERAL INFORMATION:

; APPLICANT: Dotson, Stanton B.

; APPLICANT: Kovalic, David K.

; APPLICANT: Liu, Jingdong

; APPLICANT: Lutfiyeva, Linda L.

; APPLICANT: McIninch, James

; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH

; FILE REFERENCE: 38-21(15300)J

; CURRENT APPLICATION NUMBER: US/10/155,881

; CURRENT FILING DATE: 2002-05-22

; NUMBER OF SEQ ID NOS: 37595

; SEQ ID NO 10047

; LENGTH: 332

; TYPE: PRT

; ORGANISM: Zea mays

US-10-155-881-10047

alignment\_scores:

Quality: 120.50 Length: 233

Ratio: 1.242 Gaps: 11

Percent Similarity: 41.631 Percent Identity: 25.751

alignment\_block:

US-09-462-480-3/rev x US-10-155-881-10047 ..

Align seg 1/1 to: US-10-155-881-10047 from: 1 to: 332

```
479 AGTCTTCGGCGCGGTGGCGGG.....AAGTC 451
      :|||
34 SerAlaProSerArgThrArgGlyArgArgGlyAlaAlaThrArgAl 50
450 TGTTGTCATTACGGAGCTCACCAGTCTCTCTTCGTCCTCCAGTCGCT 401
      :|||
50 aSerThrThrProAlaProProAlaArgThrSerAlaArgAlaGlyA 67
400 CGTCTCTTCTTCA.....CGTCTCGCGAGCGGTGCCGC 363
      :|||
67 laSerAlaAlaThrAlaAlaThrTrpArgThrProAlaSerGlyAlaGly 83
362 GCGACCAAGACCGGGCTGGTGGACCGCGGATTGCGAACCTGGCCCAT 313
      :|||
```

```
84 SerThrArgArgAlaThrAlaProSerProAlaArgThrAlaArgProAl 100
312 CGTCCCGGACCCACCGAGCGCGCCACCGTCACCGAC..... 273
      :|||
100 aAlaAlaGlySerAlaSerSerArgThrProArgThrSerCysAlaCys 117
272 .....GATCCGGCAACAGCGCGCGCGCATC 249
      :|||
117 erArgProArgSerSerSerProAlaProGlyAlaProProAlaLe 133
248 ACCGAGGGGCAACCGGCTTTTCGATCAGTGTAGACATCAGCGCGCTGC 199
      :|||
133 uproArg....SerProSerProThrThrAlaProArgSerGlyAlaArg 149
198 GGTCAACGACCCACCTCGCGCAGTGTAGCGCTCCGCGCAGCAGCGCCG 149
      :|||
149 rgSer.ArgAlaThr.SerProArgAlaSerCysAlaArgArgProA 165
148 CGCCC..... 144
      :|||
165 laProSerCysArgArgProSerArgProArgSerProArgHisCys 181
143 .....GGCTGGG.....CCTGATCCACCAGCCAG 118
      :|||
182 ArgProThrSerAlaValGlyAlaGlyArgAlaProGlyProProSerTh 198
117 CGGATGTTTCACAGCGGACTGTGTCCGAGCAGCGCCCATCTGCGCGCTT 68
      :|||
198 r.....ThrSerProArgSerAlaSerS 207
67 CCTCGTCGGTGGGTG..... 51
      :|||
207 erAlaSerAlaGlyProThrArgProArgAlaGlyGlyAlaAlaThr 223
51 ..... 51
224 LeuHisProArgSerArgThrAspArgLeuArgGlyAlaArgSerThrAl 240
50 .....CGCGCGCGGTGCGCGCCACCTGG 27
240 aCysProProArgGlyProProHisProAlaThrTrp 254
```

seq\_name: /cgn2\_6/ptodata/2/paa/US10\_NEW\_COMB.pep:US-10-155-881-10704

seq\_documentation\_block:

; Sequence 10704, Application US/10155881

; GENERAL INFORMATION:

; APPLICANT: Dotson, Stanton B.

; APPLICANT: Kovalic, David K.

; APPLICANT: Liu, Jingdong

; APPLICANT: Lutfiyeva, Linda L.

; APPLICANT: McIninch, James

; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH

; FILE REFERENCE: 38-21(15300)J

; CURRENT APPLICATION NUMBER: US/10/155,881

; CURRENT FILING DATE: 2002-05-22

; NUMBER OF SEQ ID NOS: 37595

; SEQ ID NO 10704

; LENGTH: 356

; TYPE: PRT

; ORGANISM: Zea mays

US-10-155-881-10704

alignment\_scores:

Quality: 120.50 Length: 140

Ratio: 1.746 Gaps: 8

Percent Similarity: 49.286 Percent Identity: 35.000

alignment\_block:

US-09-462-480-3/rev x US-10-155-881-10704 ..

370 GTGGCGGGCGGACAGACCC  
GGGCTGCTGGAG 339

```
|||||: : : |||
50 lySerAlaProSerThrProArgArgProArgAlaSerThrPro 66
338 CCGCGGATTCGGAACCGCTGCCATCGCTCCG .....G 304
|||||: : : |||
67 ProProSer.SerSerAlaGlyProProProProProSerProArgP 83
|||||: : : |||
303 ACCACCGGCGCGCGCCGTCACCGAGATCCGCAACACCGCGC 254
|||||: : : |||
83 roProProLeuThrAlaAlaProGlyPro.GlyProGlyArgThrArg 99
253 GCATCAGCGAGGGGCAACCGCTTTTCGATCAGCTGAGACATCAGCGGC 204
|||||: : : |||
99 gArgHisArg.....GlnArgAlaThrSerProAlaArgSerArgArg 114
203 GTGGCGGTCAACGAC.....CCACTGCGCCAGGTAG 172
|||||: : : |||
114 r9.ArgProCysSerAlaArgSerProProCysProProSerProArgTh 130
171 CGACTCCGCGCGCAGCGCCGCGCTGGGCGCTGATCCACCG 122
: : : ||| : : : |||
130 rThrProThrArgArgProProSerAlaAlaAlaTrpProSerSerArg 147
121 CCAGCGGATGTTGCAC.....ACGGGACTGGTCCGAGCAGGCC 81
: : : ||| : : : |||
147 rArgSerTrpAlaSerSerCysArgSerArgThrArgProSerThrPro 163
80 ATCTCGCGCGCTTCTCTCGGCT..... 57
|||||: : : |||
164 ProAlaAlaSerGlyThrSerSerProThrArgGlySerSerThrProSe 180
56 .....GGTTGCCCGCGCGTGGCGCCACC 30
|||||: : : |||
180 rArgArgArgArgArgArgArgGlyMetProProProProSerProSer 196
seq_name: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep.US-10-155-881-26034
seq_documentation_block:
; Sequence 26034, Application US/10155881
; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Lutfiyya, Linda L.
; APPLICANT: McIninch, James
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TITLE OF INVENTION: TRANSCRIPTION IN PLANTS
; FILE REFERENCE: 38-21(15300)J
; CURRENT APPLICATION NUMBER: US/10/155,881
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 37595
; SEQ ID NO 26034
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-155-881-26034
```

```
alignment_scores:
Quality: 118.50 Length: 180
Ratio: 1.362 Gaps: 7
Percent Similarity: 48.333 Percent Identity: 28.889
alignment_block:
US-09-462-480-3 x US-10-155-881-26034 ..
Align seg 1/1 to: US-10-155-881-26034 from: 1 to: 349
30 GGTGGCGCGCACCGCGCGCGCAACCGCGAGCGAGGAGGAGCGCGCAGA 79
|||||: : : |||
62 GlyGlyArgGlyGlyGluArgHisValArgArgAla...GlyG1 77
80 TGGGCTCTCTCGGCACCGATCGCTGTGCAACCATCCGCTGGCTGGTGA 129
```

```
||| |||: : : ||| : : : ||| |||
77 yGlyGlyAlaGluGluGluAlaArgArgGlyAlaGlyGlySerTrp 94
130 TCAGGCCCGCGCGCGCGCGCTCTGCGCGCGAGTGCCTACCTGG 179
|||: : : ||| |||: : : |||
94 rArgGlyArgArgGlyGlyGlyAlaGlySerProAlaAlaGlyArg 110
180 CGCAGGTGGTCTGTTACCGCGCGCGCTGATGTCTCAGTCGATCGAAA 229
|||||: : : ||| : : : |||
111 ArgArg.....ProProArgAlaProProCysAsnAlaGlyGlyG1 124
230 AGCGGTGCCCCCTCGCTGATGCGCGCGCTGTTGCGGATCGTCGCT. 278
: |||: : : ||| : : : |||
124 u.....AspValGlyGlyAlaCysGlyAlaGluGlyG 135
279 .....GACGGTGGCGCGCTCC.....GGTGGTCCGG 308
|||||: : : ||| : : : |||
135 luAlaArgGluAlaAspGlyGlyHisArgGlyGlyArgGlyGlyGluAla 151
309 AGCATGGCGCGCGGTTCGCAATCCGCGCGCTCCACCGCGCGGTCTGG 358
: : : ||| |||: : : ||| : : : |||
152 AlaGluGlyGlnGlyArgGlyAspArgAlaAspArgSerProHisLeuG1 168
359 TCGC.....GCCGGCACCGCTC 375
| yAlaArgGlyAlaProLeuGlyHisValArgArgGlyAlaGlyValAla 185
376 GCGCAGGAGCGTGAAGA..... 392
|||: : : |||
185 rProArgAlaValGlnArgGlyHisArgGlnArgProProArgArgAla 201
393 .....AGACGACGAGGACGACTGGG 412
|||||: : : |||
202 ArgAlaArgProArgArgProArgAlaProArgArgArgProArgArg 218
413 ACGAAGAGGAGGACTGGTGAGCTCCCGTAAATGACAACAGA 452
|||||: : : ||| : : : |||
218 gArgArgProArgValLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 231
```



OM of: US-09-462-480-3 to: PIR\_71:\* out\_format : pfs  
 Date: Jul 22, 2002 1:26 AM  
 About: Results were produced by the GenCore software, version 4.5,  
 Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame2\_n2p.model -DEV=xlh  
 -O=/sqn2.1/USPTO\_spool/US09462480/runat\_18072002.164418.19445/app\_query\_fasta\_1.2850  
 -DB=PIR\_71 -OFT=fastan -SUFFIX=rpr -GAPOP=12.000 -CAPEXT=4.000  
 -MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000 -XGAPOP=4.500  
 -XGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -XGAPOP=6.000  
 -DELEX=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
 -DELEX=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi  
 -LIST=45 -DSCALE=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0  
 -ALIGN=15 -MODEL=LOCAL -OUTFT=pfs -NORM=ext -HEAPSIZ=500  
 -MINLEN=0 -MAXLEN=2000000000 -USER=US09462480\_@CGN1\_1\_73  
 -NCPU=6 -ICPU=3 -LONGLOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30  
 -NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-462-480-3  
 Query length: 481  
 Database: PIR\_71:\*  
 Database sequences: 283138  
 Database length: 96089334  
 Search time (sec): 87.620000

score\_list:

Sequence	Strd	Orig	ZScore	Escore	Len	Documentation
PIR2:G70802	+	713.00	943.67	4.9e-45	368	! Probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
PIR2:T10033	+	198.00	264.03	4.3e-07	302	! Hypothetical protein MLCB28.14
PIR2:S52996	-	140.00	185.24	0.0078	403	! prpL2 protein - human (fragment)
PIR2:S20590	-	133.00	170.07	0.0242	913	! exo-alpha-sialidase (EC 3.2.1.1)
PIR2:A46302	-	131.00	169.28	0.0346	707	! PTB-associated splicing factor,
PIR1:RAMS	+	127.00	162.57	0.0672	860	! elastin precursor - mouse
PIR2:D96711	-	125.50	161.99	0.0879	708	! Hypothetical protein F24J5.8 [A
PIR2:JC4163	+	125.50	160.89	0.0869	825	! DNA-binding protein SE5 - rat
PIR2:T31611	+	125.00	155.51	0.0902	1585	! Hypothetical protein Y5088A.g
PIR2:T22602	-	124.50	166.78	0.1108	304	! Hypothetical protein F54B11.2 -
PIR2:A70507	-	124.50	159.08	0.1025	882	! Probable reductase (EC 1.3.99.1)
PIR2:T35192	-	124.00	159.65	0.1129	744	! Probable ABC transporter - Stre
PIR2:T70355	-	123.00	166.91	0.1460	227	! Hypothetical protein Rv1158c -
PIR2:S20590	+	122.50	156.19	0.1435	913	! exo-alpha-sialidase (EC 3.2.1.1)
PIR2:S51939	-	121.00	159.50	0.1953	439	! chitinase (EC 3.2.1.14) precurs
PIR2:T50568	-	120.50	150.80	0.1959	1334	! Probable multi-domain regulat
PIR2:A61183	-	119.50	159.98	0.2582	312	! Hypothetical protein (sdsB reg
PIR1:WMBE38	-	119.00	158.32	0.2782	358	! Infected cell protein ICP34.5 -
PIR2:T75457	-	119.00	157.61	0.2762	395	! Hypothetical protein - Deinoco
PIR2:S06733	-	119.00	154.36	0.2673	620	! Hydroxyproline-rich glycoprote
PIR1:GCB025	+	118.50	156.83	0.3003	402	! collagen alpha 2(I) chain - bov
PIR2:S61534	+	118.50	152.98	0.2888	684	! Abi substrate ena (enabled) - h
PIR2:T75311	+	118.00	153.87	0.3194	552	! ABC transporter, ATP-binding pr
PIR2:T404859	-	118.00	150.85	0.3097	862	! Extensin homolog F28A21.80 - Ar
PIR2:T46289	-	118.00	150.65	0.3091	862	! Hypothetical protein DKFZ0434A1
PIR2:G86441	-	118.00	148.26	0.3017	1201	! unknown protein [imported] - A
PIR2:RA1724	-	118.00	148.18	0.3015	1213	! limb deformity (ld) protein -
PIR2:T70766	-	117.50	151.28	0.3409	721	! Hypothetical protein Rv2082 - M
PIR2:RA40505	+	117.50	144.06	0.3169	1958	! Hypothetical protein - build he
PIR2:T33663	-	117.00	154.46	0.3858	424	! Hypothetical protein W04B5.3 -
PIR2:T01696	-	117.00	154.42	0.3857	426	! pistil extensin-like protein pr
PIR1:A45344	-	117.00	145.59	0.3527	1446	! immediate-early protein - suid
PIR2:T20807	+	116.50	155.39	0.4268	340	! Hypothetical protein F13A7.1 -
PIR2:S39369	-	116.50	152.01	0.4124	543	! homeotic protein BarH1 - fruit
PIR2:T43776	-	116.00	150.54	0.4453	607	! drebrin E2 - chicken
PIR2:S31521	+	115.50	147.78	0.4745	812	! collagen COL1 - freshwater sp
PIR2:S49151	+	115.50	145.03	0.4614	1188	! extensin-like protein - maize
PIR1:BA5344	+	115.50	142.30	0.4489	1733	! probable nuclear antigen - sui
PIR2:S21961	-	115.00	150.14	0.5325	534	! proline-rich protein APG - Arab
PIR2:T43556	-	115.00	149.62	0.5297	574	! Wiskott-Aldrich syndrome protei

PIR2:T38819 - 115.00 149.62 0.5297 574 ! Wiskott-Aldrich syndrome pr  
 PIR2:S54986 - 115.00 145.76 0.5094 980 ! regulatory protein - Emerice  
 PIR2:T35474 - 114.50 150.62 0.5864 456 ! 50kD proline rich protein -  
 PIR2:SL3367 - 114.50 148.57 0.5743 606 ! Om(1D) protein - fruit fly (f  
 PIR2:T13078 - 114.50 146.82 0.5642 772 ! KIAA0992 protein - human (fi  
 seq\_name: pir2:G70802  
 seq\_documentation\_block:  
 Probable PPE protein - Mycobacterium tuberculosis (strain H37RV)  
 C:Species: Mycobacterium tuberculosis  
 C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
 C:Accession: G70802  
 R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordo  
 ; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,  
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
 Nature 393, 537-544, 1998  
 A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete gen  
 A:Reference number: A70500; MUID:98295987  
 A:Accession: G70802  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-368 <COL>  
 A:Cross-references: GB:AL022120; GB:AL123456; NID:g3261558; PIDN:CAAL7965.1; PID:e12  
 A:Experimental source: strain H37RV  
 C:Genetics:  
 A:Gene: PPE

alignment\_scores:  
 Quality: 713.00 Length: 143  
 Ratio: 5.057 Gaps: 0  
 Percent Similarity: 98.601 Percent Identity: 96.503  
 alignment\_block:  
 US-09-462-480-3 x G70802  
 Align seg 1/1 to: G70802 from: 1 to: 368

1 CTGCAGCAGGTGACGTGTTGTTTCAGCCAGGTGGGGCCAGCCGGCGCGG 50  
 |||||  
 226 LeuGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGlyG 242  
 51 CAACCCAGCCGACGAGGAGCGCGCAGATGGCTGCTCGGCACCGAGTC 100  
 |||||  
 242 YAsnProAlaAspGluAlaGlnMetGlyLeuLeuGlyThrSerP 259  
 101 CGCTGTCCAAACATCCGCTGGCTGGATCAGGCCCGCCAGCGCGCGCG 150  
 |||||  
 259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275  
 151 GGCCTGCTGCGCGGAGTGCTACCTGGCGCAGGTGGGTGGTTCGTTGACCCG 200  
 |||||  
 276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292  
 201 CACGCCGCTGATGCTCAGCTGATCGAAAGCCGTTGCCCGCTCGGTGA 250  
 |||||  
 292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309  
 251 TGCCGGCGGCTGTTGCCGGATCGCTGGTACGGGTGGCGCGCGCTCCCGTG 300  
 |||||  
 309 etProAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325  
 301 GGTCCGGGAGCGGCGGCGGAGGTTCGCAATCCCGCGGCTCCACCGCCC 350  
 |||||  
 326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342  
 351 GGGTCTGTCGCGCGCGCACCGCTCGCGCAGGAGCGTGAAGAACGACGACG 400  
 |||||  
 342 oglyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspaspG 359  
 401 AGGACGACTGGGACGAAGAGACGACTGG 429



A:Residues: 1-913 <HEI>  
A:Cross-references: EMBL:X62276; NID:g39254; PIDN:CAA44166.1; PID:g39255  
C:Keywords: glycosidase; hydrolase

alignment\_scores:  
Quality: 133.00 Length: 188  
Ratio: 1.529 Gaps: 9  
Percent Similarity: 46.277 Percent Identity: 29.787

alignment\_block:

US-09-462-480-3/rev x S20590 ..

Align seg 1/1 to: S20590 from: 1 to: 913

```
472 CGCGCGGGTGGCGGAGTCTGTCTATTACGGGAGCTCACCAGTCG 423
||||| ||| ||||| ||||| ||||| ||||| ||||| |||||
586 ArgProArgProArgArgAlaLeuSerProArgHisArgHisH1 702
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
422 TCCTCTTCGTCAGTGGTCTGCTCTTCTTTCAGCTCTCGCGGAG 373
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
702 sProProArgProSerArgAlaLeuArg.....ProSerArgA 715
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
372 CGGTGCGGCGG.....CGACAGACCGGGCTGGTGAGC 338
|| ||||| ||||| ||||| ||||| ||||| ||||| |||||
715 laGlyProGlyAlaGlyAlaHisAspArgSerGluHisGlyAlaHisThr 731
|| ||||| ||||| ||||| ||||| ||||| ||||| |||||
337 CGCGGATTGCGAACCTGCGCCATCTCCCGGACCCACCGAGCGCG 288
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
732 GlySerCysAlaGlnSerAlaProGluGlnThrAspGlyProThrAlaA1 748
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
287 CCACCCGTCACCGACGTCGCGAACAGCGCGGATCACCGAGGGGCG 238
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
748 a....ProAlaProGluThrSerSerAlaProAlaAlaGluProThrClnA 764
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
237 AACGGGCTTTTCA.....TCAGCTGAG 215
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
764 laProThrValAlaProSerValGluProThrGlnAlaProGlyAlaGln 780
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
214 ACATCAGCGCG.....TCGGGTCAACGACCCACCTGCGCCAGGTAGC 171
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
781 ProSerSerAlaProLysProGlyAlaThrGlyArgAlaProSerValva 797
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
170 GACTCCGCGCGCAGCAGC..... 152
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
797 lAsnProLysAlaThrGlyAlaAlaThrGluProGlyThrProSerSers 814
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
151 .....CGCGCGCGCGGTGGGC 134
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
814 erAlaSerProAlaProSerArgAsnAlaAlaProThrProLysProGly 830
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
133 CTT.....ATCCACCGCCAGCGGTGTTTCAGACGCGGACTG 96
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
831 MetGluProAspGluLeuAspArgProSerAspGlyThrMetAlaGlnPr 847
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
95 GTGCCGAGCAGGCCATCTCGCGGCTCTCTCGTCCGCTGGGTGCGGCC 46
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
847 oThrGlyAla....ProAlaArgArgValProArgArg.....ArgA 860
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
45 GCGGTCGCCGCCA 32
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
860 rgArgArgArgPro 864
```

seq\_name: pir2:A46302

seq\_documentation\_block:

Pro-associated splicing factor, long form - human  
N:Alternate names: myoblast cell surface antigen 24.1D5; polypyrimidine tract-binding pr  
C:Species: Homo sapiens (man)  
C:Date: 21-Sep-1993 #sequence revision 18-Nov-1994 #text\_change 24-Sep-1999  
C:Accession: A46302; A43557; S29995  
R:Patton, J.G.; Porro, E.B.; Gaiceran, J.; Tempst, P.; Nadal-Ginard, B.  
Genes Dev. 7, 393-406, 1993

A:Title: Cloning and Characterization of PSF, a novel pre-mRNA splicing factor.  
A:Reference number: A46302; MUID:93194059  
A:Accession: A46302  
A:Molecule type: mRNA  
A:Residues: 1-707 <PAT>  
A:Cross-references: EMBL:X70944; NID:g38457; PIDN:CAA50283.1; PID:g38458  
A:Note: sequence extracted from NCBI backbone (NCBIF:127206)  
R:Gover, H.J.; Moore, S.E.; Dickson, G.; Elsom, V.L.; Nayak, R.; Walsh, F.S.  
Development 105, 723-731, 1989  
A:Title: Cloning and characterization of a myoblast cell surface antigen defined by 2

A:Reference number: A43557; MUID:90091812

A:Accession: A43557

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 312-707 <GOW>

A:Cross-references: GB:X16850; NID:g23711; PIDN:CAA34747.1; PID:g23712

C:Genetics:

A:Gene: GDB:SFPQ; PSF

A:Cross-references: GDB:I38275

A:Map position: 4q-4q

C:Superfamily: unassigned ribonucleoprotein repeat-containing proteins; ribonucleopro

C:Keywords: alternative splicing; pre-mRNA splicing; surface antigen

F:298-359/Domain: ribonucleoprotein repeat homology <RRM1>

F:372-438/Domain: ribonucleoprotein repeat homology <RRM2>

alignment\_scores:

Quality: 131.00 Length: 139

Ratio: 1.899 Gaps: 9

Percent Similarity: 49.640 Percent Identity: 31.655

alignment\_block:

US-09-462-480-3/rev x A46302 ..

Align seg 1/1 to: A46302 from: 1 to: 707

```
367 CGCGCGGACGACGCGGGCTGTGGAGCGCGGATTCGACACCTGG 318
||||| ||| ||||| ||||| ||||| ||||| ||||| |||||
46 ProMetGlyProGlyProGlyGln.....SerGlyProLy 57
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
317 CCATCGCTCCGACCCACCG.....GACGCGGCGCCACCGTCAC 277
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
57 sProProileProProProProHisGlnGlnGlnGlnProProp 74
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
276 CGACGATCCGCGAACACGCGCGGCATCA.....CCGAGGGGCAACCG 233
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
74 ro.....GlnGlnProProGlnGlnProProHisGlnPro 87
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
232 GCTTTTCGATCAGTACATCAGCGGCGGTGCGGTCAAGACCCACCT 183
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
88 ProProHisProGlnPro.HisGlnGln.....GlnGln.ProPro 100
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
182 CGCCAGGTAGGACTCCGCGCG.....AGCAGGCC 151
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
101 ProProProGlnAspSerSerLysProValAlaGlnGlyProGlyPr 117
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
150 CGCGCGCGCTGGGCTGATCCACGACGCGGCA..... 114
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
117 oAlaProGlyValGlySerAlaProAlaSerSerSerAlaProAla 134
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
113 .....TGGTTCGACAGCGGACTGGTG 93
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
134 laThrProThrSerGlyAlaProProGlySerGlyProGlyProThr 150
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
92 CGAGGAGGCGCATCTCGCGGGTTCCTCGTCGGCTGGTTCGCGCGCC 43
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
151 ProThrProProAlaValThrSerAlaProProGlyAlaProProPr 167
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
42 GGTGCCGCCACCC 30
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
167 oThrProProSer 171
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

seq\_name: pir1:EAMS





183 Ser.....ProProSerProProSerGluArgProThrGlnSerPro.. 196

54 GTTGGCGCGCGGTGGCGCCAC 30

197 .....ProProSerProProSer 203

seq\_name: pir2:JC4163

seq\_documentation\_block:

DNA-binding protein 5E5 - rat

C:Species: Rattus norvegicus (Norway rat)

C>Date: 27-Aug-1995 #sequence\_revision 19-Oct-1995 #text\_change 20-Jun-2000

C:Accession: JC4163; PC4040

R:Biochem. 118, 122-128, 1995

J. Suzuki, E.; Kojima, N.; Yoshimura, K.; Uyemura, K.; Obata, K.; Akagawa, K.

A:Title: Cloning and sequence analysis of cDNA for a possible DNA-binding protein 5E5 in

A:Reference number: JC4163; MUID:96015159

A:Accession: JC4163

A:Molecule type: mRNA

A:Residues: 1-825 <SU2>

A:Cross-references: DDBJ:D37934; NID:g531260; PIDN:BAA07153.1; PID:g531261

A:Experimental source: brain

A:Accession: PC4040

A:Molecule type: protein

A:Residues: 230-455 <SU2>

C:Comment: This protein has an abundance of arginine, a glycine-rich region and a prolin

C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;

C:Keywords: nerve; phosphoprotein

F:436-443/Region: nuclear location signal

F:722-731/Region: proline cluster

F:62,258,345,360,404,413,570,613,635,752,820/Binding site: phosphate (Ser) (covalent) (b

alignment\_scores:

Quality: 125.50 Length: 198

Ratio: 1.294 Gaps: 10

Percent Similarity: 48.990 Percent Identity: 31.818

alignment\_block:

US-09-462-480-3 x JC4163 ..

Align seg 1/1 to: JC4163 from: 1 to: 825

24 CAGCCAGGTGGG.....CGGCACCGCGCGGCAACCCAGCCGACGAGG 67

478 GlnAlaGlyGlyGlyProArgGlyArgAlaArgGlyProArgGlnGlnAl 494

68 AAGCGCGCAGATGGCGTGTCTGGCACCAGTCCGCTGTCGAACCATCCG 117

494 aArgArgArgHisGlyProGlnArgArgGly.....Bro.Pro 507

118 CTGGCTGGTGGATCAGCCCGCCAGCGCGCGCGCGCGCTGCTG 159

508 GlnAlaGlyGluGlyProGlyAspAlaThrLeuValLeuGlyLeuGl 524

160 .....CGCGCGGAGTCGCTACTGCGC 181

524 yThrThrSerGlyGluGlnArgAlaAspGlnSerGlnThrLeuProAlaL 541

182 CAGTGGTGTCTTGACCGC.CACGCGCGTGATGTCTACGTGATCGAATA 230

541 euAlaGlyAlaProThrAlaHisAlaHisAlaValProGlyPro 557

231 GCCGGTCCCGCCCGTGTATGTCGGC.....GGCTGTTCCCGATCGT 274

558 AlaAlaAlaThrLeuGlyGlyArgGlyArgGlySerTrpArg..... 572

275 CGGTGACGGGTGGCGCGCTCCGTGGGTCCGGAGGCGATGGCCAGGGT 324

573 .....GlyGlyArgArgGlyGlyGlyAlaGlyAlaSerGlyGlyAla 587

325 TCGCAATCCG.....CGGTCCACCGCCCGGTCTGTGGCGGCC 365

587 rgGlyGlyArgGlyArgGlyArgGlyGlyArgGlySerGlyLeuSer 603

366 GGCACC.....GCTCGCGCAGGAGCGTGAAG 391

604 GlyThrArgGluAspAlaGlySerProSerAlaArgGlyGluGlnAr 620

392 AAGACGA.....C 399

620 gArgArgGlyHisGlyProProAlaAlaGlyAlaAlaGlnValSerThrA 637

400 GAGCAGCAGTGGGAGAGAGAGAGAGTGGTGGTCCGCTAATGAC.... 446

637 rgGlyArgAlaArgGlyGlnArgThrGlyGluGluAlaGlnAspGly 653

447 .....AACAGACTTCCCGCCACCGCGGC 470

654 LeuLeuProArgGlyArgAspArgLeuProLeuArgProGly 667

seq\_name: pir2:T31611

seq\_documentation\_block:

hypothetical protein Y50E8A.g - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999

C:Accession: T31611

R:Steward, C.

A:Reference number: Z21047

A:Accession: T31611

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-1595 <WIL>

A:Cross-references: EMBL:AL117200; NID:el549770; PIDN:CAB55050.1; CESP:Y50E8A.g

A:Experimental source: clone Y50E8A

C:Genetics:

A:Gene: CESP:Y50E8A.g

A:Introns: 25/3; 60/1; 133/2; 217/3; 270/3; 337/2; 400/1; 746/2

alignment\_scores:

Quality: 125.00 Length: 157

Ratio: 1.437 Gaps: 6

Percent Similarity: 55.414 Percent Identity: 31.210

alignment\_block:

US-09-462-480-3/rev x T31611 ..

Align seg 1/1 to: T31611 from: 1 to: 1585

477 TCTTCCGCGCGGTGGCGGGAAGTCTGTTCATTACGGGAGCTCACC 428

882 AlaAlaGlyGlyGlyGlySerSer.....GlyGlyTyr. 893

427 AGTCGTCTCTTCGTCCTCCAGTCGCTCGTCTTCACGCTCCTGC 378

894 ..SerGlyGlySerSerGlyGlyGlyGlySerSerGlyGlyTyr 909

377 GCGAGCGGTGCGCGCGCCAGACCGCGGCTGGTGGAGCCCGGATTG 328

910 SerGlyGlySerAlaAlaLeuProProProProProProProProPr 926

327 CGAACCTCGGCTCGCTCCCGGACCCAGCCGCGGCGCCACCGTCA 278

926 oProProProProAlaProAlaProAlaProAlaProAlaProSerGlyG 943

277 CCGACGATCCGGCAACACCGCGCGCATCACCGAGGGGGCAACCGGCTTT 228

943 lyTyrSerGlyAlaSerSerGlyGlySerAlaAlaGlyGlyGlyGly 959

227 TCGATCAGCTGAGCATCAGCGGCGGTGGGTCAACGACCCACCTGCGCC 178

960 Ser...SerGlyGlyTyrSerGlyGlySerAlaAlaProProProProPr 975

\_\_\_\_\_

[illegible]

```

386 SerProSerAlaProAlaProGlyProSerGluProAlaSerGly..... 400
108 CGACACGGGACTGTGTCGGACGACGGCCCATCTGCGGGCTTCCTCGTCGG 59
401 .ProSerAlaProAlaProGlyProProAlaProAlaAlaGlyProSerA 417
58 CTGGGTGCGCGCGCGGTGCGCGCC 33
417 laProAlaProGlyProSerAlaPro 425

```

seq documentation block:

C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence revision 17-Jul-1998 #text change 20-Jun-2001

Citation: 17 J. Great Lakes Res. 537-544, 1991  
 Cdate: 17 J. Great Lakes Res. 537-544, 1991  
 Caccession: G70555  
 R/coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, R.; Connor, R.; Davies, R.; Devlin, K.; Fellwell, T.; Gentles, S.; Hamilton, R.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squar, Nature 393, 537-544, 1998

A:Reference number: A70500: MUID:98295987  
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the  
 A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell  
 A:Reference number: A70500: MUID:98295987

A:Accession: G70555

A: Status: preliminary: nucleic acid sequence not shown: translation not shown: 07033

A; Molecule type: DNA

A;Residues: 1-227 <COL>

A;Cross-references: GB:295584; GB:AL123456; NID:g3261774; PIDN:CAB09017.

A; Experimental source: strain H37Rv

C;Genetics:

A;Gene: Rv1158c

C;Superfamily: proline-rich protein 3

alignment_scores:		
Quality:	123.00	Length: 220
Ratio:	1.281	Gaps: 11
Percent Similarity:	43.636	Percent Identity: 26.364

alignment\_block:  
US-09-462-480-3/rev x G70555

Align seg 1/1 to: G70555 from: 1 to: 227

15 valGlySerAlaAlaLeuThrGlyGlyIleAlaHisAlaAspPr 31

428 ..... CAGTCGTCCTCTTCGT 413

[illegible]

31 oAlaProAlaProAlaProAsnIleProGlnGlnLeuIleSerS 48

[illegible]

412 CCCAGTCGTCCTCGTCTTCTTACGCTCCTGCGAGCGGTGCCGC 363

[illegible]

48 erAlaAlaAsnAlaProGlnIleLeuGlnAsnLeuAlaThrAlaLeuGly 64

362 GCGACCAGACCCGGGCTGGTGAGCCGCCGATTGCCAACCCGGCCCAT 313

65 AlaThr...ProProLeuSerAlaProLysValAlaGluProAlaProAl 80

312 CGTCCCGACCCACCGAGCGGCCACCGAC...GATCCGG 266

20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1041 1042 1043 1044 1045 1046 1047 1048 1049 1050

80 aAlaProGlyIleThrAlaThrPheProGlyLeuThrProAlaAlaProA 97

CCC

265 CAACAGCCGCC..... 233

[illegible]

СІС РІВНІСТЬ ІНТЕРІОРІАЛІВ ПІДПІСІВ / 6

254 GGCATCACCGACGGCGCAACCGGCTTT 228

234 .....GGCAACCGAGGGGGCAACCGGCII..... 228

1114 proTleproGlyVtIeThrProAlaAalaproAlaleuprovalThralapr 130

[illegible]



```
58 rProArgProProPro.....ProArgProProThrProArgp 71
315 CATCGCTCCCGGACCCACCGAGCGGCCACCGTCCACCGACGATCCGG 266
   ||| ||||| ||| ||| ||||| |||
71 roProProProArgProProThrProArgProProProProThrProArg 87
265 CAACAGCGCGCGGATCACCAGGCGGCAACCGGCTTTTCGATCAGCTGA 216
   ||||| ||| :|||
88 ProProProProArgProProThrProArgPro..... 98
215 GACATCAGCGCGTGGCGGTCAAGCAC.....CCACCTGCGCGAGGTAGCG 170
   ||| ||||| |||
99 .....ProProProProThrProArgProProProProArgProp 112
169 ACTCCGCGCGCAGCAGCGCGCGCGCTGGGGCCTGATCCACCAAGCC 120
   :|| ||| |||: ||| |||||
112 roThrProArgProProProProProThrProArgProProProPro 128
119 AGCGGATGGTTCGACAGCGGACTGGTCCGAGCAGGCCCATCTCGCGGCG 70
   :|| |||||
129 Thr.....ProArgPro..... 132
69 TTCCTCGTCGGCTGGGTGGCGCGCGGTGCGGCCACC 30
   ||||| |||||
133 .....ProProProSerProProThr 139
```

THIS PAGE BLANK (uspto)

OM of: US-09-462-480-3 to: SwissProt\_40:\* out\_format : pfs  
Date: Jul 22, 2002 1:41 AM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:  
-MODEL=frame+n2p\_model -DEV=xlh  
-O=/cgn2.1/USPNO\_spool/US09462480/runat\_18072002\_164420\_19600/app\_query.fasta\_1.2850  
-DB=SwissProt\_40 -QFWT=fastan -SUFFIX=isp -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-OGAPOP=4.000 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFWT=pfs  
-NORM-ext -HEARSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09462480 -CGN1\_1\_29 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

Search information block:  
Query: US-09-462-480-3  
Query length: 481  
Database: SwissProt\_40:\*  
Database sequences: 105224  
Database length: 38719550  
Search time (sec): 49.760000

score_list:	Sequence	Strd Orig	zScore	Escore	Len	Documentation
SwissProt_40:WASP_HUMAN	-	145.50	160.09	0.0638	503	! Q43516 homo sapiens (human)
SwissProt_40:SPPO_HUMAN	-	131.00	142.68	0.4232	707	! P23246 homo sapiens (human)
SwissProt_40:GPI_CHRE	-	130.50	143.78	0.4682	555	! Q91966 chlamydomonas reinhardtii
SwissProt_40:ELS_MOUSE	-	127.00	137.19	0.7031	860	! P54320 mus musculus (mouse)
SwissProt_40:SES_RAT	-	125.50	135.90	0.8644	825	! Q63003 rattus norvegicus (rat)
SwissProt_40:DIAL_HUMAN	-	120.50	127.92	1.59	1248	! Q60610 homo sapiens (human)
SwissProt_40:EXTN_TOBAC	-	119.00	131.03	2.15	620	! P13983 nicotiana tabacum (c
SwissProt_40:FMN_CHICK	-	118.00	125.50	2.23	1213	! Q05858 gallus gallus (chicken)
SwissProt_40:YK82_MYCTU	-	117.50	128.46	2.57	721	! Q10690 mycobacterium tuberculosis
SwissProt_40:EXLP_TOBAC	-	117.00	131.46	2.96	426	! Q03211 nicotiana tabacum (c
SwissProt_40:IE18_PRVKA	-	117.00	123.28	2.49	1446	! P33479 pseudorabies virus (c
SwissProt_40:TPM5_DROME	-	115.50	128.77	3.33	504	! P49456 drosophila melanogaster
SwissProt_40:AFSK_STRGR	-	115.50	125.62	3.31	807	! P54742 streptomyces griseus
SwissProt_40:VNUA_PRVKA	-	115.50	120.50	2.97	1733	! P33485 pseudorabies virus (c
SwissProt_40:APG_ARATH	-	115.00	127.86	3.75	534	! P40602 arabidopsis thaliana (c
SwissProt_40:T2D3_HUMAN	-	115.00	123.13	3.39	1083	! Q00268 homo sapiens (human)
SwissProt_40:HMID_DROAN	-	114.50	126.49	3.94	606	! P22544 drosophila melanogaster
SwissProt_40:SEPA_EMENI	-	114.50	119.24	3.38	1790	! P78621 emeritella nidulans
SwissProt_40:MYCN_WARMO	-	114.00	127.81	4.38	460	! Q61976 marmota monax (woodch
SwissProt_40:IE18_PRVIF	-	114.00	120.08	3.72	1461	! P11675 pseudorabies virus (c
SwissProt_40:Y735_STRFR	-	113.50	129.16	4.87	348	! P20186 streptomyces fradiae
SwissProt_40:KLF2_HUMAN	-	113.50	129.03	4.85	355	! Q95533 homo sapiens (human)
SwissProt_40:CA21_RANCA	-	113.50	120.06	4.02	1355	! Q42350 rana catesbeiana (bu
SwissProt_40:TEGU_HSV11	-	113.50	114.38	3.56	3164	! P10220 herpes simplex virus
SwissProt_40:HK61_RAT	-	112.50	127.80	5.53	365	! Q35762 rattus norvegicus (rat)
SwissProt_40:WASP_MOUSE	-	112.50	125.43	5.26	520	! P70315 mus musculus (mouse)
SwissProt_40:CAPU_DROME	-	112.50	120.67	4.75	1059	! Q24120 drosophila melanogaster
SwissProt_40:DRPL_RAT	-	112.50	119.92	4.68	1183	! P54258 rattus norvegicus (rat)
SwissProt_40:CA21_BOVIN	-	112.50	118.97	4.59	1364	! P02465 bos taurus (bovine)
SwissProt_40:D4DR_HUMAN	-	112.00	125.63	5.71	467	! P21917 homo sapiens (human)
SwissProt_40:SYNL_MOUSE	-	112.00	123.21	5.42	670	! Q88935 mus musculus (mouse)
SwissProt_40:SYNL_HUMAN	-	112.00	122.87	5.38	705	! P17600 homo sapiens (human)
SwissProt_40:NKCL_MOUSE	-	112.00	119.28	4.99	1205	! P55012 mus musculus (mouse)
SwissProt_40:ICP3_HSVIF	-	111.50	128.95	6.62	263	! P08353 herpes simplex virus
SwissProt_40:NO20_MEDTR	-	111.50	128.82	6.60	268	! P93329 medicago truncatula
SwissProt_40:EBN1_EBV	-	111.50	122.98	5.83	641	! P03211 Epstein-Barr virus (str
SwissProt_40:SYNL_BOVIN	-	111.50	122.34	5.76	706	! P17599 bos taurus (bovine)
SwissProt_40:CSP_PLASI	-	111.00	125.86	6.70	386	! Q03110 plasmodium simium. cir
SwissProt_40:Y70K_TYMCV	-	111.00	122.60	6.26	628	! P28478 turnip yellow mosaic
SwissProt_40:CCAB_RABIT	-	111.00	113.80	5.20	2339	! Q05152 oryctolagus cuniculu

SwissProt_40:ICP3_HSV11	-	110.50	128.30	7.63	248	! P36313 herpes simplex virus		
SwissProt_40:YEN1_SCHPO	-	110.50	123.14	6.84	536	! O13695 schizosaccharomyces		
SwissProt_40:NCR2_HUMAN	-	110.50	112.78	5.50	2517	! Q9Y618 h nuclear recepto		
SwissProt_40:CSP_PLAKU	-	110.00	125.45	7.77	351	! P04922 plasmodium knowlesi		
SwissProt_40:HK61_MESAU	-	110.00	125.21	7.73	364	! Q60554 mesocricetus aurat		
seq_name: SwissProt_40:WASP_HUMAN								
seq_documentation_block:								
ID	WASP_HUMAN	STANDARD;	PRT;	503 AA.				
AC	Q43516; Q9UNP1; Q15220;							
DT	16-OCT-2001 (Rel. 40, Created)							
DT	16-OCT-2001 (Rel. 40, Last sequence update)							
DT	16-OCT-2001 (Rel. 40, Last annotation update)							
DE	Wiskott-Aldrich syndrome protein interacting protein (WASP interacting							
DE	protein) (PRPL-2 protein).							
GN	WASPIP OR WIP.							
OS	Homo sapiens (Human).							
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;							
OC	Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.							
OX	NCBI_TaxID=9606;							
RN	[1]							
RP	SEQUENCE FROM N.A.							
RX	MEDLINE=98070810; PubMed=9405671;							
RA	Ramesh N., Anton I.M., Hartwig J.H., Geha R.S.;							
RT	"WIP, a protein associated with Wiskott-Aldrich syndrome protein,							
RT	induces actin polymerization and redistribution in lymphoid cells.,"							
RL	Proc. Natl. Acad. Sci. U.S.A. 94:14671-14676(1997).							
RN	[2]							
RP	SEQUENCE FROM N.A.							
RC	TISSUE=Tonsil;							
RA	Kreideweiss S., Delany-Heiken P., Nordheim A., Ruhlmann A.;							
RL	Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.							
RN	[3]							
RP	SEQUENCE OF 192-503 FROM N.A., AND VARIANT ALA-495.							
RX	MEDLINE=99218549; PubMed=10202051;							
RA	Stewart D.M., Tian L., Nelson D.L.;							
RT	"Mutations that cause the Wiskott-Aldrich syndrome impair the							
RT	interaction of Wiskott-Aldrich syndrome protein (WASP) with WASP							
RT	interacting protein.,"							
RL	J. Immunol. 162:5019-5024(1999).							
CC	-!- FUNCTION: MAY HAVE DIRECT ACTIVITY ON THE ACTIN CYTOSKELETON.							
CC	-!- INDUCES ACTIN POLYMERIZATION AND REDISTRIBUTION.							
CC	-!- SUBUNIT: BINDS TO WASP, PROFILIN AND ACTIN.							
CC	-!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN PERIPHERAL BLOOD							
CC	MONONUCLEAR CELLS, SPLEEN, PLACENTA, SMALL INTESTIN, COLON,							
CC	THYMUS. LOWER EXPRESSION IN OVARY, HEART, BRAIN, LUNG, LIVER,							
CC	SKELETAL MUSCLE, KIDNEY, PANCREAS, PROSTATE, TESTIS.							
-----								
This SWISS-PROT entry is copyright. It is produced through a collaboration								
between the Swiss Institute of Bioinformatics and the EMBL outstation -								
the European Bioinformatics Institute. There are no restrictions on its								
use by non-profit institutions as long as its content is in no way								
modified and this statement is not removed. Usage by and for commercial								
entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>								
or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).								
-----								
DR	EMBL; AF031588; AAC03767.1; -							
DR	EMBL; AF106062; RAD45972.1; -							
DR	EMBL; X86019; CAA60014.1; -							
DR	MIM; 602357; -							
DR	InterPro; IPR003124; WH2.							
DR	Pfam; PF02205; WH2; 1.							
DR	SMART; SM00246; WH2; 1.							
KW	Actin-binding; Repeat.							
FT	DOMAIN 2 13	POLY-PRO.						
FT	DOMAIN 64 96	GLY-RICH.						
FT	DOMAIN 241 244	POLY-SER.						
FT	DOMAIN 264 433	PRO-RICH.						
FT	REPEAT 352 361	XRSGFPXPX MOTIF 1.						
FT	REPEAT 374 383	XRSGFPXPX MOTIF 2.						
FT	REPEAT 410 419	XRSGFPXPX MOTIF 3.						
FT	SITE 45 48	ACTIN BINDING.						

```
FT VARIANT 495 495 G -> A.
FT CONFLICT 303 309 PHRPHLR -> SSQAPP (IN REF. 3).
FT CONFLICT 360 360 P -> PV (IN REF. 2).
FT CONFLICT 487 503 GSNRRRGGPLPIPR -> EYECOGF (IN REF. 2).
SQ SEQUENCE 503 AA; 51489 MW; 43EB88674DD3BF1A CRC64;

alignment_scores:
  Quality: 145.50      Length: 163
  Ratio: 2.021         Gaps: 10
  Percent Similarity: 44.172      Percent Identity: 33.742

alignment_block:
US-09-462-480-3/rev x WAIP_HUMAN ..
Align seg 1/1 to: WAIP_HUMAN from: 1 to: 503

472 CGGCGCGGGTGGCGGGAAGTCTGTGTACGGAGCTCACCAGTCG 423
|||||
315 ArgProGlyProPro.....
422 TCCCTTCGTCCTCCAGTCGT.....CCTC 400
|||||
320 .ProLeuProProSerSerSerGlyAsnAspGluThrProArgLeuProG 336
|||||
399 GTCGTCTTCTTTCACGCTCCTCGCGAGCGGTGCGCGCGCCAGCAGACCG 350
|||||
336 InArgAsnLeu...SerLeuSerSerThrProProLeuProSerPro 351
|||||
349 GCGTGTGGAGCCCGCGGATTGCGAACCTGCGCCATCGCTCCCGGACCC 300
|||||
352 Gly.....ArgSerGlyPro...LeuProProPr 360
|||||
299 ACCGAGCGGCGCCAGCGTCACCGAGCATCCGCGCAACAGCCCGCGCAT 250
|||||
360 oProSerGluArg...ProProProValArgSerProGlyArgS 376
|||||
249 CA.....CCGAGGGGCAACCGGCTTTCATCAGCTGAGACATCA 209
|||||
376 erGlyProLeuProProProProValSerArgAsnGlySerThrSer 392
|||||
208 CGCGCGTGGGTCAAGACCCACCTCGGCA.....GGTAGGCAC 168
|||||
393 ArgAlaLeuProAlaThr...ProGlnLeuProSerArgSerGlyValAsp 408
|||||
167 TCCGCGCGCAGCAGCGCGCGCGCGCTGCGGCTGATCCACCGACGAC 118
|||||
409 SerProArgSerGlyProArgProLeuProProAspArgPro..... 423
|||||
117 CGGATGTTGACAGCGGAGTGTGTCGCGAGCAGCGCCCATCTGCGCGGCTT 68
|||||
423 .....
67 CCTCGTGGGTGGTGGCGCGCGCGGTGGCGGCCACCC 30
|||||
424 .....SerAlaGlyAlaProProProProProSer 434
|||||

seq_name: SwissProt_40:SFPO_HUMAN

seq_documentation_block:
ID SFPO_HUMAN STANDARD; PRT; 707 AA.
AC P23246; P30808;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Splicing factor, proline- and glutamine-rich (polypyrimidine tract-
DE binding protein-associated splicing factor) (PTB-associated splicing
DE factor) (PSF) (DNA-binding P52/P100 complex, 100 kDa subunit).
GN SFPO OR PSF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND ALTERNATIVE SPLICING.
RC TISSUE=Fetal brain;
RX MEDLINE=93194059; PubMed=8449401;
RA Patton J.G., Porro E.B., Galceran J., Tempst P., Nadal-Ginard B.;
RT "Cloning and characterization of PSF, a novel pre-mRNA splicing
RT factor.";
RL Genes Dev. 7:393-406(1993).
RN [2]
RP SEQUENCE OF 312-707 FROM N.A.
RC TISSUE=Fetal skeletal muscle;
RX MEDLINE=90091812; PubMed=2480877;
RA Gower H.J., Moore S.E., Dickson G., Elsom V.L., Nayak R., Walsh F.S.;
RT "Cloning and characterization of a myoblast cell surface antigen
RT defined by 24.1D5 monoclonal antibody.";
RL Development 105:723-731(1989).
RN [3]
RP SEQUENCE OF 48-68 AND 213-246.
RX MEDLINE=93176127; PubMed=8439294;
RA Zhang W.-W., Zhang L.-X., Busch R.K., Farres J., Busch H.;
RT "Purification and characterization of a DNA-binding heterodimer of 52
RT and 100 kDa from HeLa cells.";
RL Biochem. J. 290:267-272(1993).
CC -!- FUNCTION: ESSENTIAL PRE-MRNA SPLICING FACTOR REQUIRED EARLY IN
CC SPLICEOSOME FORMATION. BINDS TO THE MAMMALIAN POLYPYRIMIDINE
CC TRACTS. FORMS A COMPLEX WITH THE POLYPYRIMIDINE TRACT-BINDING
CC PROTEIN (PTB). SEEMS TO ALSO BIND DNA.
CC -!- SUBUNIT: HETEROTETRAMER OF TWO 52 KDA AND TWO 100 KDA SUBUNITS.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; A LONG FORM (SHOWN
CC HERE) AND A SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- SIMILARITY: CONTAINS 2 RNA RECOGNITION MOTIFS (RRM).
CC -!- CAUTION: WAS ORIGINALLY (REF.2) THOUGHT TO BE MYOBLAST CELL
CC SURFACE ANTIGEN 24.1D5 AND A POSSIBLE MEMBRANE-BOUND PROTEIN
CC ECTOKINASE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X70944; CAA50283.1; -
DR EMBL; X16850; CAA34747.1; -
DR PIR; A43557; A43557.
DR PIR; S29770; S29770.
DR HSP; P19339; 1SXL.
DR MIM; 605199; -
DR InterPro; IPR000504; RRM.
DR Pfam; PF00076; rrm; 2.
DR SMART; SM00360; RRM; 2.
DR PROSITE; PS0102; RRM; 2.
DR PROSITE; PS00303; RRM_RNP_1; 1.
KW Nuclear protein; RNA-binding; DNA-binding; mRNA splicing; Repeat;
KW Alternative splicing.
FT DOMAIN 297 369 RNA-BINDING (RRM) 1.
FT DOMAIN 371 452 RNA-BINDING (RRM) 2.
FT DOMAIN 9 27 3 X 3 AA REPEATS OF R-G-G.
FT REPEAT 9 11 1.
FT REPEAT 19 21 2.
FT REPEAT 25 27 3.
FT DOMAIN 10 266 GLN/GLU/PRO-RICH.
FT DOMAIN 10 15 POLY-GLY.
FT DOMAIN 20 27 POLY-GLY.
FT DOMAIN 56 65 POLY-PRO.
FT DOMAIN 67 71 POLY-GLN.
FT DOMAIN 95 98 POLY-GLN.
FT DOMAIN 99 103 POLY-PRO.
FT DOMAIN 184 188 POLY-PRO.
FT DOMAIN 571 574 POLY-ARG.
```



219 CTGAGACATCAGCGGGTGGGGTCAACGA

219 CTGAGACATCAGCGGCGTGCGGGTCAACGACCCACCTGCCCGCAGGTAGCG 170

```
130 r. . . . . ||||| ||||| |||||
169 ACTCGCGCGCAGCAGCGCGCGCGCTGGGCTGATCCA...CCA 123
137 roSerProProSer...ProAlaProProSerProAlaPro 152
122 GCCAGCGGATGTTTCGACAGCGGAGTGGTGGCGAGCGCCACTCTGCGC 73
153 ProSer...ProSerProProValProPro 161
72 GCCTTCCTCGCTGGCTGGTGGCGCGCGCTGGCGCGCCACC 30
161 oSerProSer...ProProValProProSer 170
```

seq\_name: SwissProt\_40:ELS\_MOUSE

```
seq_documentation_block:
ID ELS_MOUSE STANDARD; PRT; 860 AA.
AC P54320;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Elastin precursor (Tropoelastin).
GN ELN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C; TISSUE=Lung;
RX MEDLINE=95130069; PubMed=7829060;
RA Wyder K.S., Sechler J.L., Boyd C.D., Passmore H.C.;
RT "Use of an intron polymorphism to localize the tropoelastin gene to
RT mouse chromosome 5 in a region of linkage conservation with human
RT chromosome 7."
RL Genomics 23:125-131(1994).
CC -!- FUNCTION: MAJOR STRUCTURAL PROTEIN OF TISSUES SUCH AS AORTA AND
CC -!- NUCAL LIGAMENT, WHICH MUST EXPAND RAPIDLY AND RECOVER COMPLETELY.
CC -!- SUBUNIT: THE POLYMERIC ELASTIN CHAINS ARE CROSS-LINKED TOGETHER
CC INTO AN EXTENSIBLE 3D NETWORK.
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX OF ELASTIC FIBERS.
CC -!- PTM: THE CROSSLINKS ARE MADE OF DEAMINATED LYS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U08210; AAA80155.1; -
DR HSSP; P04002; LWFA.
DR MGD; MGI:95317; Eln.
DR InterPro; IP0003979; tropoelastin.
DR PRINTS; PR01500; TROPOELASTIN.
KW Structural protein; Repeat; Signal; Connective tissue.
FT SIGNAL 1 27 POTENTIAL.
FT CHAIN 28 860 ELASTIN.
FT SEQUENCE 860 AA; 71955 MW; 0C0B5AAE1EDD7F1 CRC64;
```

alignment\_scores:

```
Quality: 127.00 Length: 141
Ratio: 1.649 Gaps: 8
Percent Similarity: 54.610 Percent Identity: 34.752
```

alignment\_block:

US-09-462-480-3 x ELS\_MOUSE ..

Align seg 1/1 to: ELS\_MOUSE from: 1 to: 860

```
31 GTGGCGCGCACCGGC.....GGCGCAACCC 56
:||||| |||
53 ileGlyGlyLeuGlyGlyGlyAlaLeuGlyProGlyGlyLysPr 69
57 AGCCGAGAGAAAGCCGCGAGATGGGCTCTCGGCACACGTCGGTGT 106
| ||||| :||| :|||:||||| :
69 oProLysProGlyAlaGlyLeuLeuGlyThrPheGlyAlaGlyProGlyG 86
107 CGAACCATCCGCTGGCTGGATCAGCCCGAGCGGGCGGGCGGCGCTG 156
: : |||||:|||||:||||| |||
86 ly.....LeuGlyGlyAlaGlyProGlyAlaGlyLeuGlyAla 98
157 CTGCGCGCGGAGTCTACCTCGCGCGAGTGGTGGTGTGACCCG..... 200
: : ||| :|||:|||||:||||| :
99 PheProAlaGlyThrPheProGlyAlaGlyAlaLeuVal...ProGlyGly 114
201 .....CACGCGCTGATCTCAGCTGATCGAAGCGGTTGGCC 241
||||| :||| |||||:|||||
115 AlaAlaGlyAlaAlaAlaAlaAlaAlaAlaAlaAlaGlyAlaGly 131
242 CCTCGGTGATCGCGCGCGCTGTCGCGGA.....TCGTCGGTG 279
||||| :|||:||||| |||
131 yLeuGlyGlyValGlyGlyValProGlyGlyValGlyValGlyVal 147
280 ACGGTGGCGCGCGCTCGGTGGGT...CCGGAGCGGATGGCGGAGGTC 326
|||||:||||| ||| :|||:||||| |||
148 ProGlyGlyValGlyValGlyValGlyValProGlyGlyValGlyValGly 164
327 GCATCCGCGCGC.....TCCACCA 346
: ||||| :|||:|||||
164 yValProGlyGlyValGlyGlyIleGlyGlyIleGlyGlyLeuGlyVal 181
347 GCCCGGTGCTGTCGCGCGCG 366
|| ||| |||:|||||
181 erThrGlyAlaValPro 187

seq_name: SwissProt_40:5E5_RAT
seq_documentation_block:
ID 5E5_RAT STANDARD; PRT; 825 AA.
AC Q63003;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE 5E5 antigen.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar; TISSUE=Brain;
RX MEDLINE=96015159; PubMed=8537300;
RA Suzuki E., Kojima N., Yoshimura K., Uyemura K., Obata K., Akagawa K.;
RT "Cloning and sequence analysis of cDNA for a possible DNA-binding
RT protein 5E5 in the nervous system."
RL J. Biochem. 118:122-128(1995).
CC -!- FUNCTION: MIGHT HAVE DNA-BINDING ABILITY.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN NEURONS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D37934; BAA07153.1; -
KW DNA-binding; Nuclear protein; Antigen.
SQ SEQUENCE 825 AA; 86831 MW; AF667FE2FD555BDF CRC64;
```

## alignment\_scores:

Quality: 125.50 Length: 198  
Ratio: 1.294 Gaps: 10  
Percent Similarity: 48.990 Percent Identity: 31.818

## alignment\_block:

US-09-462-480-3 x 5E5\_RAT ..

Align seg 1/1 to: 5E5\_RAT from: 1 to: 825

```
24 CAGCCAGGTGGG.....CGGCACCGCGCGCGGCAACCCACGACGACGAGG 67
Gln | | | | | | | | | | | | | | | | | | | | | | | | | | | |
478 GlnAlaGlyGlyGlyProArgGlyArgAlaArgGlyProArgGlyGlnAla 494
68 AAGCCGCGCAGATGGGCTGCTCGGCACACAGTCCGCTGTCGAACCATCCG 117
: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
494 aArgArgArgHisGlyProGlnArgArgGly.....Pro.Pro 507
118 CTGGCTGGTGGATCAGCGCCAGCGCGCGCGGCGCTGCTG..... 159
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
508 GlnAlaGlyGluGluGlyProGlyAspAlaThrLeuValLeuGlyLeuG 524
160 .....CGCGCGAGTCGCTACCTGGCG 181
524 yThrThrSerGlyGluGlnArgAlaAspGlnSerGlnThrLeuProAlaL 541
182 CAGTGGTGGTGTGACCGG.CAGCCGCTGATGTCTCAGCTGATCGAATA 230
541 euAlaGlyAlaProThrAlaHisAlaHisAlaValProGlyPro 557
231 GCGGTTGCCCCCTCGGTGATGCGGC.....GGCTGTTGCCGATCGT 274
558 AlaAlaAlaThrLeuGlyGlyArgGlyArgGlySerTrpArg..... 572
275 CGGTGACGGTGGCGCGCTCCGTCGGTCCGGAGCGATGGCGCAGGTT 324
573 .....GlyGlyArgArgGlyGlyGlyAlaGlyAlaSerGlyGly 587
325 TCGCAATCCGG.....CGGCTCCACGACCGCGGCTGTCGCGGCC 365
587 rgGlyGlyArgGlyArgGlyArgGlyArgGlySerGlyLeuSer 603
366 GGCACCC.....GCTCGCGCAGGAGCGGTGAAG 391
604 GlyThrArgGluAspAlaGlySerProSerAlaArgArgGlyGluGln 620
392 AAGACGA.....C 399
620 gArgGlyHisGlyProProAlaAlaGlyAlaAlaGlnValSerThrA 637
400 GAGCAGCACTGGGACGAGAGGACGACTGTGACCTCCGTAATGAC... 446
||||| | | | | | | | | | | | | | | | | | | | | | | |
637 rgGlyArgArgAlaArgGlyGlnArgThrGlyGluGluAlaGlnAspGly 653
447 .....AAGACACTTCCCGCCACCGCGGC 470
654 LeuLeuProArgGlyArgAspArgLeuProLeuArgProGly 667
```

seq\_name: SwissProt\_40: DIAL\_HUMAN

## seq\_documentation\_block:

ID DIAL\_HUMAN STANDARD; PRT; 1248 AA.  
AC O60610: Q9UC76;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Diaphanous protein homolog 1 (Diaphanous-related formin 1) (DRF1).  
GN DIAPH1 OR DIAPI.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

NCBI\_TaxID=9606;

[1]

SEQUENCE FROM N.A.

RX MEDLINE=98028756; PubMed=9360932;

RA Lynch E.D., Lee M.K., Morrow J.E., Welcsh P.L., Leon P.E., King M.-C.;

RT "Nonsyndromic deafness DFNA1 associated with mutation of a human

RL homolog of the Drosophila gene diaphanous.";

RN Science 278:1315-1318(1997).

[2]

SEQUENCE OF 218-817 FROM N.A.

RC TISSUE=Ovarian carcinoma;

RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,

RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,

RA Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo Y., Sugawara M.,

RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,

RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saifu K.,

RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,

RA Ninomiya K., Iwayanagi T.;

RT "NEDO human cDNA sequencing project";

RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.

[3]

SEQUENCE OF 727-765 AND 1121-1145.

RC TISSUE=Platelet;

RX MEDLINE=9525215; PubMed=7737110;

RA Reinhard M., Giehl K., Abel K., Haffner C., Jarchau T., Hoppe V.,

RA Jockusch B.M., Walter U.;

RT "The proline-rich focal adhesion and microfilament protein VASP is a

RL ligand for profilins.";

CC EMBO J. 14:1583-1589(1995).

CC -!- FUNCTION: BINDS TO GTP-BOUND FORM OF RHO AND TO PROFILIN. ACTS IN

CC A RHO-DEPENDENT MANNER TO RECRUIT PROFILIN TO THE MEMBRANE, WHERE

CC IT PROMOTES ACTIN POLYMERIZATION. IT IS REQUIRED FOR CYTOKINESIS,

CC STRESS FIBER FORMATION, AND TRANSCRIPTIONAL ACTIVATION OF THE

CC SERUM RESPONSE FACTOR. DER PROTEINS COUPLE RHO AND SRC TYROSINE

CC KINASE DURING SIGNALING AND THE REGULATION OF ACTIN DYNAMICS (BY

CC SIMILARITY). IN HEARING IT MAY PLAY A ROLE IN THE REGULATION OF

CC ACTIN POLYMERIZATION IN HAIR CELLS.

CC -!- TISSUE SPECIFICITY: EXPRESSED IN BRAIN, HEART, PLACENTA, LUNG,

CC KIDNEY, PANCREAS, LIVER, SKELETAL MUSCLE, AND COCHLEA.

CC -!- DOMAIN: DRFS ARE REGULATED BY INTRAMOLECULAR GRD-DAD BINDING WHERE

CC RHO-GTP ACTIVATES THE DRFS BY DISRUPTING THE GRD-DAD INTERACTION

CC (BY SIMILARITY).

CC -!- DISEASE: DEFECTS IN DIAPH1 ARE A CAUSE OF AUTOSOMAL DOMINANT

CC NONSYNDROMIC SENSORINEURAL DEAFNESS 1 (DFNA1).

CC -!- SIMILARITY: CONTAINS 1 GTPASE-BINDING DOMAIN (GBD).

CC -!- SIMILARITY: CONTAINS 1 FORMIN HOMOLOG 1 (FH1) DOMAIN.

CC -!- SIMILARITY: CONTAINS 1 FORMIN HOMOLOG 2 (FH2) DOMAIN.

CC -!- SIMILARITY: CONTAINS 1 FORMIN HOMOLOG 3 (FH3) DOMAIN.

CC -!- SIMILARITY: CONTAINS 1 DRF AUTOREGULATORY DOMAIN (DAD).

CC -!- SIMILARITY: BELONGS TO THE FORMIN HOMOLOG FAMILY. DIAPHANOUS

CC SUBFAMILY.

CC -!- CAUTION: REF 2 SEQUENCE DIFFERS FROM THAT SHOWN IN THAT IT SEEMS

CC TO INCLUDE INTRONIC SEQUENCE.

CC -!- DATABASE: NAMB-Hereditary hearing loss homepage;

CC NOTE=Gene page;

CC WWW="http://www.uia.ac.be/dnalab/hht/hhgenes.html".

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))

CC EMBL; AF051782; AAC05373.1; -

CC EMBL; AK023345; BAB14533.1; ALT\_SEQ.

CC MIM; 602121; -

CC MIM; 124900; -

CC InterPro: IPR003104; FH2.

CC Pfam: PF02181; FH2; 1.

CC SMART; SM00498; FH2; 1.

CC Coiled coil; Repeat; Deafness.

CC

DR

DR

DR

DR

DR

DR

DR

DR

KW



415 hrTyrSerProProProThrTyrAlaGlnProProProLeuPro... 430  
 173 AGCGACTCCGCGCGCAGCAGCGCGCGCGCTGGGGCTGATCCACC 124  
 431 .....ProThrTyrSerProProProProAlaTyrSerProProPr 445  
 123 AGCCAGCGGATGTTTCACAGCGAGCTGGTGCAGCAGCGCCATCTGCG 74  
 445 oProThr.....TyrSerProProProProThrTyrS 456  
 73 CGGCTTCCTCGTGGCT...GGTTCGCGCGCGCGTGGCGCCACCTGG 27  
 456 eProProProProAlaTyrAlaGlnProProProProProThrTyr 472  
 seq\_name: SwissProt\_40:FMN\_CHICK

seq\_documentation\_block:  
 ID FMN\_CHICK STANDARD; PRT; 1213 AA.

AC Q05858;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Formin (Limb deformity protein).  
 GN LD.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauromia; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN=WHITE LEHORN; TISSUE=Embryo;  
 RX MEDLINE=92112031; PubMed=1730407;  
 RA Trumpf A., Blundell P.A., de la Pompa J.L., Zeller R.:  
 RT "The chicken limb deformity gene encodes nuclear proteins expressed  
 in specific cell types during morphogenesis.";  
 RL Genes Dev. 6:14-28(1992).  
 CC -!- FUNCTION: IS IMPORTANT FOR MORPHOGENESIS OF LIMB AND KIDNEY AND  
 MAY BE INVOLVED IN DETERMINING DORSOVENTRAL NEURAL TUBE POLARITY  
 AND MOTOR NEURON INDUCTION. IT MAY ALSO HAVE A FUNCTION IN  
 DIFFERENTIATED CELLS OR BE INVOLVED IN MAINTAINING SPECIFIC  
 DIFFERENTIATED STATES.  
 CC -!- SUBCELLULAR LOCATION: Nuclear.  
 CC -!- ALTERNATIVE PRODUCTS: MANY DIFFERENT FORMS OF THIS PROTEIN ARE  
 PRODUCED BY ALTERNATIVE SPLICING OF THE LD GENE. A VARIATION  
 IN SPLICING IS SEEN AMONG DIFFERENT TISSUES AND DIFFERENT SIZE  
 TRANSCRIPTS EXIST WITHIN ANY ONE TISSUE.  
 CC -!- TISSUE SPECIFICITY: PRESENT IN THE ADULT BRAIN, KIDNEY,  
 BRAIN, HEART AND INTESTINE AND THROUGHOUT THE EMBRYO.  
 CC -!- DEVELOPMENTAL STAGE: IN THE DEVELOPING LIMB BUD, THE PROTEIN IS  
 EXPRESSED IN THE APICAL ECTODERMAL RIDGE AND THE MESENCHYMAL  
 COMPARTMENT, PREDOMINANTLY IN THE POSTERIOR REGION. DURING  
 KIDNEY MORPHOGENESIS, EXPRESSION IS INITIALLY RESTRICTED TO  
 THE EPITHELIAL COMPARTMENT OF THE PRONEPHROS AND MESONEPHROS.  
 CC -!- SIMILARITY: CONTAINS 1 FORMIN HOMOLGY 1 (FH1) DOMAIN.  
 CC -!- SIMILARITY: CONTAINS 1 FORMIN HOMOLGY 2 (FH2) DOMAIN.  
 CC -!- SIMILARITY: BELONGS TO THE FORMIN HOMOLGY FAMILY. CAPPUCCINO  
 SUBFAMILY.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----

EMBL; X62681; CAA44555.1;  
 InterPro; IPR003104; FH2.  
 DR InterPro; IPR001265; Formin.  
 DR Pfam; PF02181; FH2; 1.  
 DR PRINTS; PR00828; FORMIN.  
 DR SMART; SM00498; FH2; 1.

KW Nuclear protein; Developmental protein; Coiled coil;  
 KW Alternative splicing.  
 FT DOMAIN 428 450 COILED COIL (POTENTIAL).  
 FT DOMAIN 503 572 COILED COIL (POTENTIAL).  
 FT DOMAIN 652 751 FHI (PRO-RICH).  
 FT DOMAIN 766 1171 FH2.  
 FT DOMAIN 1050 1125 COILED COIL (POTENTIAL).  
 SQ SEQUENCE 1213 AA; 135240 MW; ADE3EF0B3FB9D862 CRC64;

alignment\_scores:  
 Quality: 118.00 Length: 107  
 Ratio: 2.511 Gaps: 5  
 Percent Similarity: 43.925 Percent Identity: 31.776

alignment\_block:

US-09-462-480-3/rev x FMN\_CHICK ..

Align seg 1/1 to: FMN\_CHICK from: 1 to: 1213

353 CCCGGGCTGGTGGAGCGCGCGGATTCGGAACCTGGCCCATCGTCCCG 304  
 678 ProGlyLeuValProProPro.....ProLeuProThrGl 690  
 303 ACCACCGGAGCGCGCGCCGCTACCGGAGCTCCGCGCAACAGCCGCG 254  
 690 yProThrSerValThrProHisPheAlaPheGlyProProLeuProPro 707  
 253 GCATCACCGAGGGGCAACCGCGCTTTTCGATCAGCTGAGACATCAGCGG 204  
 707 InLeuSerGluGlyCysArgAspPhe..... 715  
 203 GTGCGGTCAACGACCCACCTGCGCGCAGGTAGCGACTCCGCGCGCAGCAG 154  
 716 ...GlnAlaProAlaProAlaPro.....Pr 724  
 153 GCCCGCGCGCGCTGGGCGCTGATCCACGAGCAGCGGATGTTGACACA 104  
 724 oProLeuProGlyLeuGlyProProValPro..... 734  
 103 CGCGACTGTTGCGGAGCAGCGCCATTCGCGGCTTCCTGCTGCGGCTGGG 54  
 735 .....ProProLeuPro.....GlySerGly 741  
 53 TTGCGCGCGCGGTGCGCGCC 33  
 742 LeuProProProProPro 748

seq\_name: SwissProt\_40:YK82\_MYCTU

seq\_documentation\_block:

ID YK82\_MYCTU STANDARD; PRT; 721 AA.

AC Q10690;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hypothetical 73.6 kDa protein Rv2082.

GN Rv2082 OR MTCv49.21.

OS Mycobacterium tuberculosis.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Corynebacteriaceae; Mycobacterium.

OX NCBI\_TaxID=1773;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=H37RV;

RX MEDLINE=98295987; PubMed=9634230;

RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,

RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekai F.,

RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,

RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,

RA Horsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,

RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,

RA Rutter S., Seeger K., Skellton S., Squares S., Squares R.,

RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
RL complete genome sequence.";  
RL Nature 393:537-544(1998).  
CC -1- SIMILARITY: SOME, TO M.TUBERCULOSIS RV0029.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; 273966; CAA98194.1; -  
DR Tuberculinist; RV2082; -  
KW Hypothetical protein; Complete proteome.  
FT DOMAIN 295 307 ALA-RICH.  
SQ SEQUENCE 721 AA; 73564 MW; 28CAC21029ED0C57 CRG64;  
  
alignment\_scores:  
Quality: 117.50 Length: 184  
Ratio: 1.366 Gaps: 7  
Percent Similarity: 46.739 Percent Identity: 25.000  
  
alignment\_block:  
US-09-462-480-3/rev x YK82\_MYCTU ..  
Align seg 1/1 to: YK82\_MYCTU from: 1 to: 721  
  
473 CCGCCCGCGG.....GTGCGCGGGAAGTCTGTGTTCATTACGGG 436  
|||||  
184 ProAlaArgProArgProValThrAlaGluAspMetThrSerMetThrAl 200  
|||||  
435 AGCTACACAGTCGTCCTGTCGCCAGTCTGCTCTGTCGTCTTCTTCAC 386  
: |||||  
200 aAnSerProAlaGlySerProPheGlyAlaAlaProSerAlaProSerH 217  
385 GCTCCTCGCGGAGCGGTGCGCGCGACACGCCGCGCTGGTGGAGCGG 336  
|||||  
217 isSerThrThrSerGlyProProThrAlaProThrProThrSerPro 233  
|||||  
335 CCGGATTGCGAACCTCGCCATCGCTCCGCGACCCACCGAGCGCGGCC 286  
|||||  
234 PheGlyThrAlaProMetValLeuSerSerSerThrSerSerGlyPr 250  
285 ACCGCTACCGACATCCGCGACACGCGCGCGGATCACCAGGGGGCAA 236  
|||||  
250 oProThrAlaProThrProThrSerProPheGlyThrAlaProMetProp 267  
235 CCGGCTTTTCGATCATCGATGAGACATCAGCGCGTGGCGGTCAACGACCA 186  
|||||  
267 roGly.....Pro 269  
185 CCGCGCCAGTACGACTCGCGCGCAGCAGCGCC...GGCGCCGGCT 139  
|||||  
270 ProProGlyThrValSerProProLeuProProSerAlaProAlaVa 286  
138 G.....GGGCTGATCCACGACCGGATGG..... 111  
: |||||  
286 lGlyValGlyGlyProSerValProAlaAlaGlyMetProProAlaAla 303  
110 .....TTCGAC 105  
303 laAlaAlaThrAlaProLeuSerProGlnSerLeuGlyGlnSerPheThr 319  
104 AGCGGACTGTGCGAGCAGGCCCATCTGCGCGGCTTCCTCG.....TC 61  
320 ThrGlyMetThrThrGlyThrProAlaAlaAlaGlyAlaGlnAlaLeuSe 336  
60 GGCTGGG.....TTCGCGCGCGCGTGGCGG 35  
|||||

336 rAlaGlyAlaLeuHisAlaAlaThrGluProLeuProProProAlaProp 353  
34 CC 33  
||  
353 ro 353  
  
seq\_name: SwissProt\_40:EXLP\_TOBAC  
seq\_documentation\_block:  
ID EXLP\_TOBAC STANDARD; PRT; 426 AA.  
AC Q03211;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE Pistil-specific extensin-like protein precursor (PELP).  
OS Nicotiana Tabacum (Common tobacco).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.  
OX NCBI\_TaxID=4097;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV PETITE HAVANA; TISSUE=Pistil;  
RX MEDLINE=93005740; PubMed=1392607;  
RA Goldman M.H., Pezzotti M., Seurinck J., Mariani C.;  
RT "Developmental expression of tobacco pistil-specific genes encoding  
RT novel extensin-like proteins.";  
RL Plant Cell 4:1041-1051(1992).  
CC -1- TISSUE SPECIFICITY: PISTIL (STIGMA AND STYLE TISSUE).  
CC -1- DEVELOPMENTAL STAGE: EXPRESSION BEGINS IN FLORAL BUDS AFTER  
CC PISTIL DIFFERENTIATION AND LEVELS GRADUALLY INCREASE DURING  
CC FLOWER DEVELOPMENT TOWARD ANTHESIS. LEVELS GRADUALLY DECREASE  
CC AFTER POLLINATION AND ARE ABSENT BY THE SIXTH DAY AFTER  
CC POLLINATION.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; Z14019; CAA78397.1; -  
DR PIR; JQ1696; JQ1696.  
DR InterPro: IPR000419; Pollen\_Ole\_e\_1.  
DR Pfam: PF01190; Pollen\_Ole\_e\_1;  
KW Structural protein; Signal; Repeat; Glycoprotein.  
FT SIGNAL 1 23  
FT CHAIN 24 426 PISTIL-SPECIFIC EXTENSIN-LIKE PROTEIN.  
FT DOMAIN 69 182 4 X 5 AA REPEATS OF S-P(4).  
FT REPEAT 69 73 1.  
FT REPEAT 76 80 2.  
FT REPEAT 83 87 3.  
FT REPEAT 178 182 4.  
FT CARBOHYD 310 310 N-LINKED (GLCNAC...) (POTENTIAL).  
SQ SEQUENCE 426 AA; 44278 MW; 51A495CC94017812 CRG64;  
  
alignment\_scores:  
Quality: 117.00 Length: 149  
Ratio: 1.625 Gaps: 7  
Percent Similarity: 48.322 Percent Identity: 28.859  
  
alignment\_block:  
US-09-462-480-3/rev x EXLP\_TOBAC ..  
Align seg 1/1 to: EXLP\_TOBAC from: 1 to: 426  
  
472 CGGCCGGGTGGCGGGAAGTCTGTGTTCATTACGGGAGCTCACCAGTCG 423  
: |||||  
138 LysProSerSerProSerProLeuValLysProProProProProSe 154

seq\_name: SwissProt\_40:IE18\_PrvKA

P33479;	
01-FEB-1994 (Rel. 28, Created)	
01-FEB-1994 (Rel. 28, Last sequence update)	
01-FEB-1994 (Rel. 28, Last annotation update)	
Immediate-early protein IE180.	
IE.	
GN	Pseudorabies virus (strain Kaplan) (PRV).
OC	Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC	Alphan herpesvirinae; Varicellovirus.
OX	NCBI_TaxId=33703;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=91021039; PubMed=2171211;
RT	Vicek C., Kozmink Z., Paces V., Schirm S., Schwzyer M.;
RA	"Pseudorabies virus immediate-early gene overlaps with
RT	oriented open reading frame: characterization of their
RT	enhancer regions.";
RL	Virology 179:365-377(1990).
CC	-I- FUNCTION: THIS IE PROTEIN IS A MULTIFUNCTIONAL PROT
CC	OF MIGRATING TO THE NUCLEUS, BINDING TO DNA, TRANS-
CC	OTHER VIRAL GENES, AND AUTOREGULATING ITS OWN SYNTH
CC	-I- SUBCELLULAR LOCATION: NUCLEUS OF INFECTED CELLS.
CC	-I- PTM: A LONG STRETCH OF SERINE RESIDUES MAY BE A MAJ
CC	PHOSPHORYLATION.
CC	-I- SIMILARITY: BELONGS TO THE ICP4/IE140/IE180 FAMILY.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced thro
CC	between the Swiss Institute of Bioinformatics and the
CC	the European Bioinformatics Institute. There are no r
CC	use by non-profit institutions as long as its conte
CC	modified and this statement is not removed. Usage by
CC	entities requires a license agreement (See <a href="http://www.i">http://www.i</a>
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
CC	-----
DR	EMBL; M34651; AAA47470.1; -
DR	PIR; A45344; A45344.
KW	Early protein; Transcription regulation; Trans-acting

KW	DNA-binding; Phosphorylation; Nuclear protein.
FT	DOMAIN 347 354 POLY-SER.
FT	DOMAIN 379 397 POLY-SER.
SQ	SEQUENCE 1446 AA; 148640 MW; 81f43a3DE3DDA068 CRC64;

  

alignment_scores:			
Quality:	117.00	Length:	185
Ratio:	1.345	Gaps:	13
Percent Similarity:	47.027	Percent Identity:	32.973

  

alignment_block:	
US-09-462-480-3/rev x IE18_PrvKA	..

  

Align seg 1/1	to: IE18_PrvKA	from: 1	to: 1446
---------------	----------------	---------	----------

  

473	CGGGCCGGTGGCCGGAGATCTGTGTCATTACGGGAGCTCACCAGTC	424	
370	ProAlaAlaAlaAlaArgArgSerAla	.....Se	379
423	GTCTCTTTCGTCCAGTCGCTCGTCTTCTTTCACGCTCTCGCGCA	374	
379	rSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer	396	
373	GGGTCGCCGGC	.....CGGACACAGCCGGGCTGTGTGAGCGC	336
396	ex-SerGluGlyGluGluAspGluGlyValArgProGly	.....Ala	409
335	CGGGATTGCAACCTCTGGCCCATCGCTCCCGGACCCACGGAGCGCGC	286	
410	Pro.LeuAlaArgAlaGlyProPro	.....ProSerPro	421
285	ACCGTCACCGCAGCATCCGGCAACAGCCGCGCATCACCGAGGGGGCA	236	
422	.....AlaProAlaAlaAlaProArg	.....	428
235	CGGGCTTTTCGATCAGCTGAGACATACGCGGCTCGGGTCAACGACCA	186	
429	ProSerAlaSerSerAlaSerSerSerAlaAlaAlaSerProAla	..Pro	444
185	..CTTGGCCAGGTAGCGACTCCGCGCGCAGCAGG	.....	153
445	AlaProGluProAlaArgProProArgArgLysArgArgSerThrAsnAs	461	
152	.....CCGCGCGCGCGCTGGGGCT	....	132
461	nHisLeuSerLeuMetAlaAspGlyProProProThrAspGlyProLeu	478	
131	.....GATCCACCGCCAGCGGA	114	
478	euThrProLeuGlyGluProTrpProGlySerAspProProAlaAspGly	494	
113	TGGTTTCGACGGCAGTGTGTCGCGAGAGGCCCATCTCGCGGGCTCCTC	64	
495	ArgValArgTyrGly	.....GlyAlaGlyAspSe	504
63	GTGCGGTGGTTCGCGCGCGGTGCGCCACCTGCTGAACAACGAGC	14	
504	rArgGluGlyLeu	.....TrpAspGluAspAspV	514
13	TC	12	
514	al	514	

  

seq_name:	SwissProt_40:TPM5_DROME
-----------	-------------------------

  

seq_documentation_block:					
ID	TPM5_DROME	STANDARD;	PRT;	504	AA.
AC	P49456;				
DT	01-FEB-1996	(Rel. 33, Created)			
DT	01-FEB-1996	(Rel. 33, Last sequence update)			
DT	30-MAY-2000	(Rel. 39, Last annotation update)			
DE	Tropomyosin 1, fusion protein 34.				





```
seq_name: SwissProt_40:VNUA_PRVKA
seq_documentation_block:
ID VNUA_PRVKA STANDARD; PRT; 1733 AA.
AC P334B5;
DT 01-FEB-1994 (Rel. 28, Created)
```

DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 01-FEB-1994 (Rel. 28, Last annotation update)  
DE Probable nuclear antigen.  
DE Pseudorabies virus (strain Kaplan) (PRV).  
OS Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Alphaherpesvirinae; Varicellovirus.  
OX NCBI\_TaxID=33703;  
RN [1]  
RP SEQUENCE FROM N. A.  
RX MEDLINE=91021039; PubMed=2171211;  
RA Vitek C., Kozmik Z., Paces V., Schirm S., Schwyzer M.;  
RT "Pseudorabies virus immediate-early gene overlaps with an oppositely  
RT oriented open reading frame: characterization of their promoter and  
RT enhancer regions."  
RL Virolgy 179:365-377(1990).

[illegible][illegible]

CC - TO SEND AN EMAIL AT THE CONCLUSION!

DR EMBL; M34031; AAA474/1.1; .  
DR PIR: B45344: B45344.

```

FT DOMAIN 112 117 POLY-TRK.
FT DOMAIN 179 173 GLY-RICH.
FT DOMAIN 192 196 POLY-SER.
FT DOMAIN 271 298 POLY-PRO.
FT DOMAIN 304 308 POLY-ARG.
FT DOMAIN 883 889 POLY-GLY.
FT DOMAIN 1398 1405 POLY-GLY.
SQ SEQUENCE 1733 AA; 172166 MW; 0C8CD8BE475BB5E2 CRC64;

alignment_scores:
    Quality: 115.50      Length: 187
    Ratio: 1.283        Gaps: 10
    Percent Similarity: 48.128      Percent Identity: 29.947

alignment_block:
US-09-462-480-3 x VNUA_PRVKA  ..

Align seg 1/1 to: VNUA_PRVKA from: 1 to: 1733

31 GTGGCGGCACCGGC.....GGCGG 50
   ||||| |||
840 ValGlyGlyGluGlyArgLeuGlyProArgValGlyLeuAlaG1 856

51 CAACCCAGCCGACGAGGACCGCGCAGATGGCGCTGCTCGGCACCAAGTC 100
   |::: ||| |||::: |||::: |||::: |||::: |||:::
856 yArgAspAlaAlaG1uAlaAlaValGlyArgGlyValLeuGlyHisGlyP 873

101 CG...CTGTCGAACCATCGCTG.....GCTGGTGGATCAGGC 135
   || ||::: |||::: |||::: |||::: |||::: |||:::
873 rogluArgAlaProgluProValValLeuGlyGlyGlyGlyGly 889

136 CCACGCGGGCGCGGGCGCTG..... 156
   ::: |||::: |||:::
890 GlnGlnArgGlySerGlyValArgSerGlyProGluSerGluGlyAlaA1 906

157 .....C 157
906 aLeuAlaProGlyProProValLeuPheValValAlaValAlaValAlav 923

```

```
158 TCGCGCGGAG.....TCGTACT 177
:: |||||
923 alProAlaGlyArgAlaGlyGluProLeuValLeuAlaValPro 939
178 GCGCAGTGGTGTGTCACCGGACGGCGGTGATCTCAGCTGATCGA 227
|||||::|||
940 GlyAlaAlaGly.....ProGlyArgAlaAlaLeuLeuLe 951
228 AAAGCGGTGCGCCCTCGGTGATCGCGGCGGTGTCGCGGATCGTCGG 277
|||||::|||
951 uAlaProLeuGlyArgTrpValArg...AlaGlyGlyGlyAlaGlyV 967
278 TCACGGTGGCGCGCTCGGTGGTCCGGGAGCG..... 312
|||||::|||
967 alAlaGlyAlaGlyGluAlaGlyLeuGlyAlaGlyAlaGlyLeuGly 983
313 ATGGCCAGGTTTCCATCCGGGGTCCAC.CAGCCCGGTCTGTGTCG 361
|||
984 AlaGlyAlaGlyLeuGlyAlaGlyAlaGlyGlyProGlyAlaGlyG 1000
362 CGCGGCGACCGCTCGCGCAGGCGGTGAAGAAGACGAGGAGGACTGG 411
|||||::|||
1000 uAlaGly.....GlyGlyAlaArgArgArgArgArgArg.Trp 1013
412 GACGAAGAG 420
|||||
1014 AspAspGlu 1016
```

seq\_name: SwissProt\_40:APG\_ARATH

seq\_documentation\_block:

```
ID APG_ARATH STANDARD; PRT; 534 AA.
AC P40602;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Anter-specific proline-rich protein APG precursor.
GN APG.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94004980; PubMed=8401599;
RA Roberts M.R., Foster G.D., Blundell R.P., Robinson S.W., Kumar A.,
RA Draper J., Scott R.;
RT "gametophytic and sporophytic expression of an anther-specific
Arabidopsis thaliana gene.";
RL Plant J. 3:111-120(1993).
CC -!- TISSUE SPECIFICITY: FOUND IN SPOROPHYTIC AND GAMETOPHYTIC CELL
CC TYPES IN THE ANTHEP, ONLY IN MALE FERTILE PLANTS.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED IN MALE GAMETOGENESIS, DURING
CC MICROSPORE DEVELOPMENT. HIGHER EXPRESSION IS FOUND DURING
CC MICROSPORE MITOSIS WITH A DRAMATIC DECLINE DURING POLLEN
CC MATURATION.
CC -!- SIMILARITY: BELONGS TO THE "GDSL" FAMILY OF LIPOLYTIC ENZYMES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X60377; CAA42925.1;
CC PIR; S21961; S21961.
CC InterPro; IPR001087; Lipase_GDSL.
CC Pfam; PF00657; Lipase_GDSL; 1.
CC PROSITE; PS01098; LIPASE_GDSL_SER; 1.
CC Signal.
```

```
FT SIGNAL 1 35
FT CHAIN 36 534 ANTER-SPECIFIC PROLINE-RICH PROTEIN APG.
FT ACT_SITE 211 211 BY SIMILARITY.
FT ACT_SITE 511 511
FT ACT_SITE 511 511
SQ SEQUENCE 534 AA; 57967 MW; 744CAD3B08CC482E CRC64;
```

alignment\_scores:

Quality: 115.00 Length: 154

Ratio: 1.691 Gaps: 9

Percent Similarity: 44.156 Percent Identity: 29.221

alignment\_block:

US-09-462-480-3/rev x APG\_ARATH

Align seg 1/1 to: APG\_ARATH from: 1 to: 534

```
463 TGGCGGGGAAGTCTGTTCATTACGGAGATCACAGTCTCTCTTCG 414
|||||
43 TrpProTrpProLeuTrpProArgProTrpProGlnProTrpProMetAs 59
413 TCCAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 364
|||||
59 nProProThrProAsp.....ProSerProLysProValAlaPro 73
363 CGCGACACAGACCCCGGCTGTGTGGAGCCGCGGATTCGGAACCCCTGCCCA 314
|||||
73 roGlyProSerPro.....LysProValAlaProProGlyPro 85
313 TCG...CTCCCGACCCACCGGAGCGGCCAC..... 284
|||||
86 SerProCysProSerProProLysProGlnProLysProProAl 102
283 .CCCTCACCAGGATCCGGCAACAGCCGCGCATCACCGAGGGGCAAC 235
|||||
102 aProSerProSerProCysProSerProProLysProGlnProLysP 119
234 CGGCTTTTCGATCAGCTGAGACATCAGCGCGTGGGTCAACAGCCAC 185
|||||
119 roVal.....ProProProAlaCysProProThr...Prop 130
184 CTGCGCCAGGTAGGACTCCGCGCGCAGCAGCGCGCGCGCGCTGGGG 135
|||
130 roLysPro.....GlnProLysProAlaProProProAlaProLys 143
134 CCTGATCCACAGCAGCGGATGTTTCGACAGCGGACTGGTCCGAGCAG 85
|||||
144 ProAlaProProProAla.....ProLys 151
84 GCCCATCTGCGCGCTTCC.....TCGTCGGCTGGTTCGCGCGC 44
|||||
151 sProValProCysProSerProProLysProProAlaProThrProLysP 168
43 CGGTGCGCGCC 33
|||||
168 roValProPro 171
```

OM of: US-09-462-480-3 to: SPTREMBL\_19:\* out\_format : pfs  
Date: Jul 22, 2002 1:40 AM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

## Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh  
-Q=/cgn2.1/USPTO.spool/US09462480/runat-18072002\_164419\_19544/app\_query.fasta\_1.2850  
-DB=SPTREMBL\_19 -QFMT=fastan -SUFFIX=rspt -GAPOB=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000  
-GAPOB=4.500 -GAPEXT=0.050 -XGAPOB=10.000 -XGAPEXT=0.500  
-FGAPOB=6.000 -FGAPEXT=7.000 -YGAPOB=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRLX=blousum62  
-TRANS=human40.cgi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09462480\_ECGN1\_1\_133 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

## Search information block:

Query: US-09-462-480-3  
Query length: 481  
Database: SPTREMBL\_19:\*  
Database sequences: 562222  
Database length: 172994929  
Search time (sec): 143.350000

## score\_list:

Sequence	Strd	Orig	zScore	EScore	Len	Documentation
sp_bacteriap:O69738	+	713.00	955.62	1.9e-45	368	O69738 mycobacterium tuberculosis
sp_bacteriap:O33085	+	198.00	266.81	5.4e-07	302	O33085 mycobacterium leprae. hy
sp_bacteriap:Q9EY22	-	145.50	197.95	0.0044	251	Q9EY22 paracoccus pantotrophus.
sp_human:Q9BU37	-	138.00	184.97	0.0164	358	Q9BU37 homo sapiens (human). si
sp_bacteriap:Q9DXH2	-	134.00	173.38	0.0341	763	Q9DXH2 mycobacterium tuberculosis
sp_bacteriap:Q59164	-	133.00	170.56	0.0408	913	Q59164 actinomycetes viscosus. si
sp_plant:Q9FPQ6	-	130.50	171.31	0.0611	555	Q9FPQ6 chlamydomonas reinhardt
sp_mammal:Q9N1P0	-	128.50	165.43	0.0880	818	Q9N1P0 bos taurus (bovine). sub
sp_invertebrate:Q9NHW4	+	128.50	157.11	0.0931	2249	Q9NHW4 nephrilia clavipes (orb
sp_virus:IP89459	-	128.50	154.41	0.0948	3122	IP89459 herpes simplex virus (b
sp_invertebrate:O44358	+	128.00	164.25	0.0963	871	O44358 nephrilia clavipes (orb
sp_rodent:Q9WUE8	-	127.50	156.39	0.1101	2087	Q9WUE8 rattus norvegicus (rat)
sp_rodent:Q9WU13	-	127.50	156.11	0.1103	2158	Q9WU13 rattus norvegicus (rat)
sp_rodent:Q9WV48	-	127.50	156.08	0.1103	2167	Q9WV48 rattus norvegicus (rat)
sp_rodent:Q9ES29	+	127.00	163.50	0.1139	810	Q9ES29 mus musculus (mouse). el
sp_plant:Q9SX31	-	125.50	162.60	0.1463	708	Q9SX31 arabidopsis thaliana (mc
sp_invertebrate:Q9F2N5	-	125.00	160.06	0.1614	889	Q9F2N5 streptomyces coelicolor
sp_invertebrate:Q20739	-	124.50	168.22	0.1658	304	Q20739 caenorhabditis elegans
sp_invertebrate:Q9VZC2	-	124.50	165.06	0.1693	446	Q9VZC2 drosophila melanogaste
sp_bacteriap:O33268	-	124.50	159.45	0.1758	882	O33268 mycobacterium tuberculosis
sp_bacteriap:O69995	-	124.00	160.18	0.1898	744	O69995 streptomyces coelicolor.
sp_bacteriap:O06556	-	123.00	168.61	0.2111	227	O06556 mycobacterium tuberculosis
sp_virus:Q9IPQ8	+	123.00	160.78	0.2225	588	Q9IPQ8 cynomolgus epstein-barr
sp_mammal:Q9N1P1	+	123.00	160.20	0.2234	631	Q9N1P1 bos taurus (bovine). sub
sp_bacteriap:Q59164	+	122.50	156.49	0.2485	913	Q59164 actinomycetes viscosus. si
sp_invertebrate:Q9VQ94	-	122.50	154.29	0.2522	1192	Q9VQ94 drosophila melanogast
sp_invertebrate:Q96016	-	122.50	154.29	0.2522	1193	Q96016 drosophila melanogast
sp_human:Q96508	-	122.00	153.68	0.2747	1184	Q96508 homo sapiens (human). s
sp_invertebrate:Q9NHW2	+	122.00	149.86	0.2818	1884	Q9NHW2 nephrilia madagascari
sp_rodent:Q88493	-	122.00	147.03	0.2873	2657	Q88493 mus musculus (mouse). t
sp_bacteriap:Q68872	-	121.50	159.44	0.2867	542	Q68872 myxococcus xanthus. hypc
sp_rodent:Q99K31	-	121.50	158.38	0.2887	616	Q99K31 mus musculus (mouse). si
sp_human:Q9Z2P0	-	121.50	153.22	0.2989	1154	Q9Z2P0 homo sapiens (human). k
sp_plant:O42421	-	121.00	160.50	0.3088	439	Q42421 beta vulgaris (sugar bee
sp_rodent:Q62775	-	121.00	160.28	0.3093	451	Q62775 rattus norvegicus (rat).
sp_virus:Q69270	-	121.00	159.96	0.3099	469	Q69270 equine herpesvirus 1. ir
sp_rodent:Q9Z0G8	-	121.00	159.68	0.3105	485	Q9Z0G8 rattus norvegicus (rat).
sp_plant:Q9FUZ7	-	121.00	157.05	0.3161	688	Q9FUZ7 zea mays (maize). scared
sp_rodent:Q9CTW8	-	120.50	163.80	0.3277	271	Q9CTW8 mus musculus (mouse). 49
sp_invertebrate:Q9BIV0	+	120.50	161.72	0.3323	349	Q9BIV0 argiope aurantia. majfd

sp\_bacteria:Q9RKR9 - 120.50 150.69 0.3579 1334 ! Q9RKR9 streptomyces coelic  
sp\_plant:Q9SI74 - 119.50 161.30 0.3922 312 ! Q9SI74 arabidopsis thaliana  
sp\_invertebrate:Q26116 + 119.50 159.74 0.3963 377 ! Q26116 plasmodium vivax.  
sp\_invertebrate:Q9BIU6 + 119.50 158.40 0.3999 444 ! Q9BIU6 argiope trifasciata  
sp\_plant:Q9FFW5 - 119.50 154.88 0.4095 681 ! Q9FFW5 arabidopsis thaliana  
seq\_name: sp\_bacteriap:O69738  
seq\_documentation\_block:  
ID O69738 PRELIMINARY; PRT; 368 AA.  
AC O69738;  
DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE PPE-FAMILY PROTEIN  
GN RV3873 OR MTV027.08.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H37RV;  
RA MEDLINE=98295987; PubMed=9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
RAC Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
Rutter S., Seeger K., Skelton S., Squares R., Squares R.,  
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
RT complete genome sequence."  
RL Nature 393:537-544 (1998).  
DR EMBL; AL022120; CAA17965.1; -.  
DR TuberculList; RV3873; -.  
DR InterPro; IPR000030; PPE.  
DR Pfam; PF00823; PPE; 1.  
KW Complete proteome.  
SQ SEQUENCE 368 AA; 37330 MW; D78F44095F658CA2 CRC64;  
alignment\_scores:  
Quality: 713.00 Length: 143  
Ratio: 5.057  
Percent Similarity: 98.601 Percent Identity: 96.503  
alignment\_block:  
US-09-462-480-3 x O69738 ..  
Align seg 1/1 to: O69738 from: 1 to: 368  
1 CTGCAGAGTGAGTGGTGTTCAGCCAGTGGCGGCGGCGGCGG 50  
|||||  
226 LeuGlnValThrSerLeuPheSerGlnValGlyThrGlyGly 242  
51 CAACCCAGCGACGACGAGGCGGCGGCGGCGGCGGCGGCGG 100  
|||||  
242 YAsnProAlaAspGluGluAlaGlnMetGlyLeuLeuGlyThrSerP 259  
101 CGCTGTGCAACCATCCGCTGGCTGGTGGATCAGGCCCGCGGCGGCGG 150  
259 rleuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275  
151 GGCTGTGTCGCGGAGTGCCTACCTACCTGGCGGCGGCGGCGGCGGCGG 200  
|||||  
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292  
201 CAGCCGCTGATGCTCTCAGCTGATCGAGAAAGCGGTTGCCCGCTCGGTGA 250  
|||||  
292 gThrProLeuMetSerGlnLeuGluLysProValAlaProSerValM 309



```

366 CGCGG.....CGACACAGACCGCGGTGGTGGAGC...CCCGCGATT 329
   ::  |||||  |||  |||  |||  |||||  ::
113 laSerArgLysAsnArgProLysProPripArgSerArgArgVal 129
   ::  |||||  |||  |||  |||  |||||  ::
328 GCGAACCTGGC.....CCATCGCTCCCGGACCCACCGGA 294
   ::  :::::  |||  ::  |||||  |||  ::
130 SerLysProSerArgArgSerArgProSerSerLeuProArgProH1 146
   ::  :::::  |||  ::  |||||  |||  ::
293 GCGGCGCCACCGCTACCGACGATCGGCAACAGCCGCGGCATCACCGA 244
   ::  |||||  |||  ::  |||||  |||  ::
146 saArgThrProArgArgHisArgMetArgArgProGlySer.... 161
   ::  :::::  |||  ::  |||||  |||  ::
243 GGGGGCAACCGGCTTTCGATCAGCTGACATCAGCGGGTGGGTCA 194
   |||  ::  ::  ::  ::  ::  ::
162 .....ArgLeuProSerLeuProSerAlaArgArgPro 172
   ::  ::  ::  ::  ::  ::
193 ACAGCCACCTGCGCGAGGTAGGACTCCGCGCGCAGCGCGCGCC 144
   |||  ::  |||  ::  |||  ::  |||||  ::
173 ThrProArgLeuLeuAsnProProArgProThrProArgLysProArgPr 189
   |||  ::  |||  |||  |||  |||  |||  ::
143 GCCTGGGGCTGATCCACGACGCGGATGTTGACAGCGGACTGGT 94
   |||  ::  |||  |||  |||  |||  |||  ::
189 oArg..AsnProLysPro.....SerGlyGlnAr 198
   |||  ::  |||  |||  |||  |||  |||  ::
93 GCGGACGAGCCCATC.....TCGCGGCTTCCTCGCGGTG 56
   |||  ::  |||  |||  |||  |||  |||  ::
198 gProLysArgProLeuHisProLysProCysArgArgProSerProSerP 215
   |||  ::  |||  |||  |||  |||  |||  ::
55 GGTTCGCGCGCGGTGCGCGCC 33
   |||  |||  |||  |||  |||  |||  |||  ::
215 roProProSerProLeuProPro 222

```

seq\_name: sp\_human:Q9BU37

```

seq_documentation_block:
ID Q9BU37 PRELIMINARY; PRT; 358 AA.
AC Q9BU37:
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE SIMILAR TO WISKOT-ALDRICH SYNDROME PROTEIN INTERACTING PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MELANOMA.;
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC002914; AAH02914.1; -.
DR InterPro; IPR002965; P_rich_extensn.
DR TrEMBL; P0003124; WH2.
DR Pfam; PF02205; WH2; 1.
DR PRINTS; PR01217; PRICHEXTENSN.
DR SMART; SM00246; WH2; 1.
SQ SEQUENCE 358 AA; 36464 MW; D008B60E60EE94EA.CRC64;

```

```

alignment_scores:
  Quality: 138.00      Length: 155
  Ratio: 2.226        Gaps: 6
  Percent Similarity: 40.000  Percent Identity: 31.613

```

alignment\_block:

```

US-09-462-480-3/rev x Q9BU37 ..
Align seg 1/1 to: Q9BU37 from: 1 to: 358

```

```

412 CCAGTGGTCTCGTCTCTTCAGCTCCGCGGAGCGGTGCGGC 363
   |||  ::  |||  ::  |||  ::  |||  ::
161 ProProGluProGlnArg.....AsnArgMetProProPr 172

```

```

362 GCGACACAGACCGCGGTGGTGGAGCCCGCGATTGCGAACCCCTGGCCCAT 313
   |||||  |||  |||  |||  |||  |||  |||  ::
172 oArgProAspValGlySerLysProAspSerIleProProValProS 189
   |||||  |||  |||  |||  |||  |||  |||  ::
312 CGTCCCGGACCCA..... 299
   ||  |||  |||
189 erThrProArgProIleGlnSerSerLeuHisAsnArgGlySerProPro 205
   |||||  |||  |||  |||  |||  |||  |||  ::
298 .....CCGGAGCGCGCCACCGCTACCGACGATCCGCGCAACA 261
   |||||  |||  |||  |||  |||  |||  |||  ::
206 ValProGlyGlyProArgGlnProSerProGlyProThrProProPr 222
   |||||  |||  |||  |||  |||  |||  |||  ::
260 GCCCGCGCATCACCGAGGGGCA.....C 235
   |||  ::  |||  |||  ::
222 oProValArgAspProProGlyArgSerGlyProLeuProProProPr 239
   |||||  |||  |||  |||  |||  |||  |||  ::
234 CGCTTTTCAGTCAGTCAGATCAGCGGCGGTGCGGTCAACGACCCAC 185
   |||||  |||||  ::  |||  |||  |||  |||  ::
239 roValSerArgAsnGlySerThrSerArgAlaLeuProAlaThr..ProG 255
   |||||  |||||  |||  |||  |||  |||  |||  ::
184 CTGCGCCA.....GGTAGGCACTCCGCGCGCAGCAGCGCGCGCC 144
   |||  |||  |||  |||  |||  |||  |||  ::
255 lnLeuProSerArgSerGlyValAspSerProArgSerGlyProArgPro 271
   |||||  |||||  |||  |||  |||  |||  |||  ::
143 GCCTGGGCGCTGATCCACGACGCGGATGTTGACAGCGGACTGGT 94
   |||  |||||  |||
272 ProLeuProAspArgPro..... 278
93 GCGGACGAGCCCATCTGCGCGGCTTCCTCGTGGTGGTGGTGGCGCGC 44
   |||||  |||  |||  |||  |||  |||  |||  ::
279 .....SerAlaGlyAlaProProp 285

```

seq\_name: sp\_bacteria:Q9XDH2

```

seq_documentation_block:
ID Q9XDH2 PRELIMINARY; PRT; 763 AA.
AC Q9XDH2:
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PROLINE-RICH MUCIN HOMOLOG.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=20090472; PubMed=10627046;
RA Espitia C.; Lacleite J.P.; Mondragon-Palomino M.; Amador A.;
RA Campuzano J.; Martens A.; Singh M.; Cicero R.; Zhang Y.; Moreno C.;
RT "The PE-PGRs glycine-rich proteins of Mycobacterium tuberculosis: a
  new family of fibronectin-binding proteins?";
RL Microbiology 145:3487-3495(1999).
DR EMBL; AF071081; RAD41594.1; -.
DR InterPro; IPR002951; Atrophin.
DR InterPro; IPR003882; Pistil_extensin.
DR InterPro; IPR002965; P_rich_extensn.
DR PRINTS; PR01222; ATROPHIN.
DR PRINTS; PR01217; PRICHEXTENSN.
DR PRINTS; PR01218; PSTLEXTENSIN.
SQ SEQUENCE 763 AA; 75035 MW; 39168EC45A5916F8.CRC64;

```

```

alignment_scores:
  Quality: 134.00      Length: 138
  Ratio: 1.887        Gaps: 8
  Percent Similarity: 51.449  Percent Identity: 36.957

```

## alignment\_block:

US-09-462-480-3/rev x Q9XDH2

Align seg 1/1 to: Q9XDH2 from: 1 to: 763

```
421 CCCTTCGTCAGTCCTCTCGTCGTCTTTCACACTCTCTCGGCGAGC 372
||||| ||| ::||| |||
551 ProProAlaProThrProLysLeuSerAlaAsnProProCy 567
||||| ||| ||| ::||| |||
371 GGTGCGCGCGAGCACACCGCGGCTGGTGGAGCGCGGATTCGGAAC 322
: ||||| ||| ||| ::||| |||
567 sProProValProProAlaPro.....AsnArgProProAlaProp 581
321 CTGGC.....CCATCGCTCCCGGACCCAGCGGAGCGGCGCAC 284
||||| ||| ::||| |||
581 roAlaProProAlaProProGluLeuProAlaProProAsp.....Pro 595
283 CGGTACCGAGGATCCCGGCAACAGCCCGCGGATCACCGAGGGGCAACC 234
||||| ||| ::||| |||
596 ProThrProProValAlaAsnSerProProAlaProProAla..... 609
233 GGCTTTTCGATCAGTCAGACATCAGCGGCTGGGTCAGACACCCACC 184
||| ||| ||| ||| |||
610 .....ProProAlaProProSerAla..LeuProPheValAsnProPr 623
183 TGGCGCAGGTAGGACTCCGCG...CGCAGCAGCGCGCG...CCGCGC 140
||||| ||| ::||| |||
623 oAlaProProThrProAlaAlaProLysSerArgProAlaLeuProAlaA 640
139 TGGGCGCTGATCCACAGCCAGCGGATGGTTCGACAGCGGATGGTCCG 90
||| ||| ||| ||| |||
640 laProProAlaProProAlaProProValArgAlaThrThrProProPro 656
89 AGCAGCGCCATCTCGCGGCTTCTCGTCGCTGGGTTGCGG.....CC 46
::: ||| ||| ||| ||| |||
657 AlaProProAlaProProAlaProAsnSerMetAlaLeuProProAlaPr 673
```

45 GCCGGTCCGCC 33

|||||

673 oProAspProPro 677

seq\_name: sp\_bacteria:Q59164

## seq\_documentation\_block:

```
ID Q59164 PRELIMINARY; PRT; 913 AA.
AC Q59164;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE SIALIDASE (EC 3.2.1.18).
GN NANH.
OS Actinomyces viscosus.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomycetes.
OX NCBI_TaxID=1656;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM43798;
RA Henningsen M.;
RL Submitted (SEP-1991) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM43798;
RA Henningsen M., Roggentin P., Schauer E.R.;
RT "Cloning sequencing and expression of the sialidase gene from
RL Biol. Chem. Hoppe-Seyler 372:1065-1072(1991).
DR EMBL; X62276; CAA44166.1;
DR HSSP; Q02834; 1EUR.
DR InterPro; IPR002860; BNR.
DR Pfam; PF02012; BNR; 5.
KW Hydrolase; Glycosidase.
```

SQ SEQUENCE 913 AA; 96216 MW; AD22CF56706FF373 CRC64;

## alignment\_scores:

Quality: 133.00 Length: 188

Ratio: 1.529 Gaps: 9

Percent Similarity: 46.277 Percent Identity: 29.787

## alignment\_block:

US-09-462-480-3/rev x Q59164

Align seg 1/1 to: Q59164 from: 1 to: 913

```
472 CGGCCCGGTGGCGGAAGTCTGTTCATTACGGAGGTCTACCACTGC 423
||||| ||| ||| ||| |||
686 ArgProArgArgProArgArgAlaLeuSerProArgHisArgHisHi 702
422 TCCTCTTCGTCGCCAGTCGCTCGCTCTTTCACGCTCTCGCGGAG 373
: ||| ||| ||| ||| |||
702 sProProArgProSerArgAlaLeuArg.....ProSerArgA 715
372 CGGTGCGCGCG.....CGACCAGACCGCGGCTGTGGAGC 338
||| ||||| ||| ||| |||
715 lagLProGlyAlaGlyAlaHisAspArgSerGluHisGlyAlaHisThr 731
337 CGCGGATTGCGAACCTCGCTCCCGGACCCACCGGAGCGCGC 288
||||| ||| ||| ||| |||
732 GlySerCysAlaGlnSerAlaProGluGlnThrAspGlyProThrAlaA 748
287 CCACCCGTCACCGACGATCGCGCAACAGCCCGCGGATCACCGAGGGGC 238
||||| ||| ||| ||| |||
748 a...ProAlaProGluThrSerSerAlaProAlaAlaGluProThrGlnA 764
237 AACCGGCTTTTCGA.....TCAGCTGAG 215
||||| ||| ||| ||| |||
764 laProThrValAlaProSerValGluProThrGlnAlaProGlyAlaGln 780
214 ACATACGCGCG.....TGCGGTCAACGACCCACCTCGCGCCAGTAGC 171
||||| ||| ||| ||| |||
781 ProSerAlaProLysProGlyAlaThrGlyArgAlaProSerValVa 797
170 GACTCGCGCGCAGCAGC..... 152
: ||||| ||| ||| ||| |||
797 lAsnProLysAlaThrGlyAlaAlaThrGluProGlyThrProSerSers 814
151 .....CGCGCGCGCGCTGGGC 134
||||| ||| ||| ||| |||
814 erAlaSerProAlaProSerArgAsnAlaAlaProThrProLysProGly 830
133 CTG.....ATCCACCGCAGCGGATGTTTCGACAGCGGACTG 96
: ||| ||| ::||| ||| |||
831 MetGluProAspGluIleAspArgProSerAspGlyThrMetAlaGlnPr 847
95 GTGCCGAGGAGGCCATCTCGCGGCTTCCTCGCTCGGCTGGGTTCGCCGC 46
||| ||||| ||| ||||| |||
847 oThrGlyAla...ProAlaArgArgValProArgArg.....ArgA 860
45 GCCGGTGGCGGCCCA 32
||||| ||||| |||
860 rgArgArgPro 864
```

seq\_name: sp\_plant:Q9FPQ6

## seq\_documentation\_block:

```
ID Q9FPQ6 PRELIMINARY; PRT; 555 AA.
AC Q9FPQ6;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE VEGETATIVE CELL WALL PROTEIN GP1.
GN GP1.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
```

OC Chlamydomonadaceae; Chlamydomonas.

OX NCBI\_TaxID=3055;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Ferris P.J., Woessner J.P., Waffenschmidt S., Kilz S., Drees J.,  
 RA Goodenough U.W.;  
 RA "Glycosylated polypyrrolone II rods-with-kinks as a structural motif in  
 RT plant hydroxyproline-rich glycoproteins.";  
 RL Biochemistry 0:0-0(2001).  
 DR EMBL; AF309494; AAG45420.1; -  
 DR InterPro; IPR003882; Pistil\_extensin.  
 DR InterPro; IPR002965; P-rich\_extensin.  
 DR PRINTS; PRO1217; PRICEXTENSIN.  
 DR PRINTS; PRO1218; PSLEXEXTENSIN.  
 SQ SEQUENCE 555 AA; 54219 MW; 6A584A90465502F5 CRC64;

alignment\_scores:  
 Quality: 130.50 Length: 148  
 Ratio: 2.008 Gaps: 9  
 Percent Similarity: 43.919 Percent Identity: 35.135

alignment\_block:  
 US-09-462-480-3/rev x Q9FPQ6 ..

Align seg 1/1 to: Q9FPQ6 from: 1 to: 555

```

469 CCGGGTGGCGGGAAGTCGTGTCATTACGGGAGTCACCAGTCGTCC 420
||||| |||||
68 ProGlyProProSer.....Pr 73
419 TCTTCGTCCACGTCGTCGTCTCTTCACGCTCTCTCGCGAGCGG 370
| ||||| |||
73 oAlaProProSerProPro.....SerProAlaProProS 85
369 TGGCGGGCGACAGACCCGGGTGGTGGAGCCGCGGATGGCAACCT 320
||||| ||||| |||||
85 erProAlaProProSerProAlaProProSerProAlaProProS 101
319 GGCCCATCGTCGCGGACCCAGCGAGCGGCCACCCGTCACCGAGAT 270
||||| ||| ||||| |||||
102 AlaProProSerProAlaProProSerProAlaProProSerProAla.. 117
269 CCGGCAACAGCCGCGCATCACCGAGGGGGCAACGGCTTTTCGATCAG 220
||||| ||| ||||| |||||
118 .....ProProSerProProSerProAlaProPro..SerProSe 130
219 CTGAGACATCAGCGGTGCGGGTCAACGACCCACCTGCGCGAGGTAGCG 170
| ||||| ||||| |||||
130 r.....ProProAlaProProSerP 137
169 ACTCCGCGCGCAGCAGCCGCGCGCGCTGGGCTGTATCCA...CCA 123
||| ||| ||||| ||| ||||| |||||
137 roSerProProSer...ProAlaProProLeuProProSerProAlaPro 152
122 GCACGCGGATGGTTCGACACGGACTGTGTCGCGAGCAGGCGCATCTCGCG 73
||| ||||| |||||
153 ProSer.....ProSerProProValProPr 161
72 GGCTTCTCTCGGTGGTGGCGCGCGCTGGGCTGTATCCA 30
::: ||| ||||| ||||| |||||
161 oSerProSer.....ProProValProProSer 170

```

seq\_name: sp\_mammal:Q9N1P0

seq\_documentation\_block:

ID Q9N1P0 PRELIMINARY; PRT; 818 AA.  
 AC Q9N1P0;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
 DE SUBMAXILLARY MUCIN (FRAGMENT).  
 GN BSM1.

OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRED ANGUS;  
 RX MEDLINE=20223253; PubMed=10759843;  
 RA Jiang W., Gupta D., Gallagher D., Davis S., Bhavanandan V.P.;  
 RT "The central domain of bovine submaxillary mucin consists of over 50  
 RT tandem repeats of 329 amino acids: chromosomal localization of the  
 RT BSM1 gene and relations to ovine and porcine counterparts.";  
 RL Eur. J. Biochem. 267:2208-2217(2000).  
 DR EMBL; AF178428; AAF67279.1; -  
 FT NON\_TER 1  
 FT NON\_TER 818  
 SQ SEQUENCE 818 AA; 73014 MW; B5BB44F84F66F86B CRC64;

alignment\_scores:  
 Quality: 128.50 Length: 182  
 Ratio: 1.397 Gaps: 9  
 Percent Similarity: 50.549 Percent Identity: 30.220

alignment\_block:  
 US-09-462-480-3/rev x Q9N1P0 ..

Align seg 1/1 to: Q9N1P0 from: 1 to: 818

```

464 GTGCCGGGAAGTCTGTGTCAATACGAGCTCACCATGTCGTCTTTC 415
::: ||||| ::||| ||||| |||
355 lIeSerGlyThrAsnValProValSerGlyAlaProValThrProGlySe 371
414 GTCCCACTGCTCTCTCG.....T 398
||||| ::||| |||||
371 rSerAlaGlySerSerGlyAlaProGlyThrGlyGlyProGlySerGluT 388
397 CGTCTTCTTCCAGCTCTGCGAGCGGTGCGGCGCAGCAGCAGCGGG 348
::: ||||| ||| ||||| ::||| ||||| |||
388 hrAlaSerProLeuSerGlyAlaAlaGlyThrSerAlaThrGlySerGly 404
347 CTGTTGGAGCCGCGGATTCGAAACCTGCGCCATCGTCCCGACCC... 300
||||| ::||| |||||
405 ThrSerIleProProSerGlyAla.....ProValThrProGluProPr 419
299 .....ACGGAGCGGG.....CCACCGCTCACCGACGATCCGG 266
||||| ||||| ::||| ||||| |||||
419 oLeuIleSerThrGlyAlaSerAlaGlyProProAlaSerSerGluSerT 436
265 CAACAGCCGCGGCATCACCGAGGGGCAACCGCTTTTCGATC..... 222
::: ||||| ||||| |||||
436 hrValThr.....LeuProGlyAlaThrGlyThrAspValLeuArg 449
221 .....AGCTGAGACATCAGCGGTGCGGTCAACACCCACCTGC 181
||| ::||| ||||| |||||
450 SerGlyThrSerLeuProValSerGlyAlaValThrProAlaProSe 466
180 GCCAGTAGCAGCTCCGCGCAGCAGGCGC..... 150
::||| ::||| |||||
466 rProGlyGlySerSerAlaThrAlaGlyProGlyValGlySerAlaThrT 483
149 .....GCGCCGCGCTGGGCTGTATCCA..... 126
::: ||| ||| |||
483 hrValGlnAlaSerGlyAlaThrGlyAlaAspValLeuArgSerGlyThr 499
125 .....CCAGCCAGCGATGTTTCGACAGC...GGACTGGTGGCCGAGCAG 85
||| ::||| ||||| |||||
500 SerLeuProValSerGlyValAlaValSerProGlySerSerProGlyAr 516
84 GCCCATCTGCGCGGCTTCCTCGTGGTGGGTGCGCGCGCGCGGTG 39
| ::||| ||||| |||||
516 gSerGlyAlaThrAlaValSerSerGlnGlySerGlnProThrVal 531

```





419 TCITCGTCCAGTCGTCTGCTGCTCTTCTTACGCTCTCTGCGGAGCGG 370  
 2710 oASpGlnProThrProGluSerAlaProProAlaTrpValSerAlaL 2727  
 369 TGCGCGCGGACACAGACCGCGGCTGGAGCCCGGATGCGAACCCCT 320  
 2727 euProLeuProGlyProAla.....SerAlaArgGlyAlaPhePro 2741  
 319 GGCCCATCGCTC.....CCGACCCACCG..... 296  
 2742 AlaProThrLeuAlaProIleProProProAlaGluGlyAlaValva 2758  
 295 .....GAGCGCGCCAC.....CCGT 280  
 2758 lProGlyGlyAspArgArgGlyArgGlnThrThraGlyProS 2775  
 279 CACGAGATCCGCGCACACGCGCGCA..... 251  
 2775 erProThrProArgGlyProAlaAlaGlyProProArgLeuThr 2791  
 250 .....TCACC 246  
 2792 ArgProAlaValAlaSerLeuSerAlaSerLeuAsnSerLeuProSerPr 2808  
 245 GAGGGGCAACCGCTTTTCGATCAGTCAGACATCAGCGCGTCCGGT 196  
 2808 oArg....AspProAlaAspHisAlaAlaAlaValSerAlaAlaAla 2824  
 195 CAAGCACCCACCTCGCCAGGT.....AGCGACTCCGCGCGCACGA 155  
 2824 laVal..ProProSerProGlyLeuAlaProProThrSerAlaValGInt 2840  
 154 GGCGCGCGCGCTGGGCTCATCA.....CCAGCC 120  
 2840 hrSerProProProLeuAlaProGlyProValAlaProSerGluProLeu 2856  
 119 AGCGAGTGGTTCGACAGCGGCTGGTCCGCGGCTCCGCGGCG 70  
 2857 CysglyTrp.....ValValPro.GlyGlyProValAlaArgA 2869  
 69 TTCCTCTCGCTGGTGGTTCGCGCGCGTCCGCGGCTCCGCGGCTGAACA 20  
 2869 rgPro.....ProProGlnSerPro 2875  
 19 ACGAGCTCACCTGCTGCA 2  
 2876 AlaThrLysProAlaAla 2881

seq\_name: sp\_invertebrate:044358

seq\_documentation\_block:  
 ID 044358 PRELIMINARY; PRT; 871 AA.  
 AC 044358;  
 DT 01-JUN-1998 (TrEMBLrel. 06, Created)  
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE FLAGELLIFORM SILK PROTEIN (FRAGMENT).  
 GN FLAG  
 OS Nephila clavipes (Orb spider).  
 OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;  
 OC Araneomorphae; Entelegynae; Araneolidae; Tetragnathidae; Nephila.  
 OX NCBI\_TaxID=6915;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98153262; PubMed=9480768;  
 RA Hayashi C.Y., Lewis R.V.;  
 RT "Evidence from flagelliform silk cDNA for the structural basis of  
 RT elasticity and modular nature of spider silks.";  
 RL J. Mol. Biol. 275:773-784(1998).  
 DR EMBL: AF027972; AAC38846.1; -.  
 DR InterPro: IPR000087; Collagen.  
 DR InterPro: IPR000209; Peptidase\_s8.

DR PROSITE; PS00138; SUBTILASE\_SER; UNKNOWN\_1.  
 FT NON TER 871 871  
 SQ SEQUENCE 871 AA; 71039 MW; 1FA1E3B7E0C5983A CRC64;

## alignment\_scores:

Quality: 128.00 Length: 132  
 Ratio: 1.707 Gaps: 6  
 Percent Similarity: 56.818 Percent Identity: 35.606

## alignment\_block:

US-09-462-480-3 x 044358 ..

Align seg 1/1 to: 044358 from: 1 to: 871

37 GGCACCGCGCGGCAACCCAGCCGACGAGGAGCCGCGCAGATGGCGCT 86  
 ||| |||||:||||:||||: |||  
 377 GlyProGlyGlyAlaGlyProGlyGlyAlaGlyProGlyGlyAlaGlyPr 393  
 87 GCTCGGCACCATCGCTGTGCAACCATCCGCTGGTGTGATCAGCGCC 136  
 ||| :|||: ||| |||||:||||  
 393 oGlyGlyTyrglyProGlyGlySerGlyPro.....GlyGlyAlaGlyP 408  
 137 CCAGCGCGCGCGCGCTGCTGCGCGGAGTCTGCTACCTGCGCGCAGGT 186  
 |||| ||||| ||||| |||||:||||  
 408 roSer...GlyAlaGlyLeuGlyGlyAlaGly.....ProGlyGlyAla 421  
 187 GGTCTGTGACCCGCGCCGCTGTGCTCAGCTGATCGTGAAGAACCGGT 236  
 |||  
 422 Gly..... 422  
 237 TGCCCGCTCGGTGATGCGCGCGGTGTTCGCGGATCGTCTGCGGCTG 286  
 :||: ||||| ||||| |||||: |||||  
 423 .....LeuGlyGlyAlaGlyProGlyGlyAlaGlyThrSerG 435  
 287 CGCGCGCTCCG.....CTGGTCCGGAGCGATGGCCAGGTTTCGCAA 330  
 :||: |||||: |||||: |||||: |||||: |||||  
 435 lAlaGlyProGlyGlyAlaGlyProGlyGlyAlaGlyGlnGlyAspAla 451  
 331 TCCGCGCGCTCCACCGCC.....GGTCTGCTGCGCGCGCGCAGCGCTCG 376  
 :||: |||||: |||||: |||||: |||||: |||||: |||||  
 452 GlyProGlyGlyAlaGlyArgGlyGlyAlaGlyArgGlyGlyValGly 468  
 377 CGCAGGAGCTGAAGAACGACGACGACGACTGGGAGCAAGAGGA 422  
 ||: ||||| ||| ||||| ||| |||||  
 468 gGlyGlyAlaGlyArgGlyGlyAlaGlyArgGlyGlyAlaArgGly 483

seq\_name: sp\_rodent:09WUE8

seq\_documentation\_block:  
 ID 09WUE8 PRELIMINARY; PRT; 2087 AA.  
 AC 09WUE8;  
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE SHANK1A.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SPRAGUE-DAWLEY;  
 RX MEDLINE=99360650; PubMed=10433268;  
 RA Naisbitt S., Kim E., Tu J.C., Xiao B., Sala C., Valtchanoff J.,  
 RA Weinberg R.J., Worley P.F., Sheng M.;  
 RT "Shank, a novel family of postsynaptic density proteins that binds to  
 RT the NMDA receptor/PSD-95/GKAP complex and cortactin.";  
 RL Neuron 23:569-582(1999).  
 DR EMBL: AF131951; AAD29417.1; -.  
 DR HSSP: P00519; IABL.  
 DR InterPro: IPR002110; ANK.  
 DR InterPro: IPR001478; PDZ.

DR InterPro: IPR002965; P-rich\_extensn.  
DR InterPro: IPR001660; SAM.  
DR InterPro: IPR001452; SH3.  
DR Pfam: PF00023; ank; 6.  
DR Pfam: PF00595; PDZ; 1.  
DR Pfam: PF00536; SAM; 1.  
DR Pfam: PF00018; SH3; 1.  
DR PRINTS; PR01217; PRICHEXTENS.  
DR SMART; SM00248; ANK; 3.  
DR SMART; SM00228; PDZ; 1.  
DR SMART; SM00454; SAM; 1.  
DR SMART; SM00326; SH3; 1.  
DR PROSITE; PS50088; ANK\_REPEAT; 3.  
DR PROSITE; PS50297; ANK\_REPEAT\_REGION; 1.  
DR PROSITE; PS50106; PDZ; 1.  
DR PROSITE; PS50002; SH3; 1.  
KW ANK repeat; Repeat.  
SQ SEQUENCE 2087 AA; 218125 MW; 0955C33D7A7AA8F8 CRC64;

alignment\_scores:  
Quality: 127.50 Length: 124  
Ratio: 1.903 Gaps: 9  
Percent Similarity: 54.032 Percent Identity: 36.290

alignment\_block:

US-09-462-480-3/rev x Q9WUE8

Align seg 1/1 to: Q9WUE8 from: 1 to: 2087

353 CCCGGG.....CTGGTGGAGCCGCCGAT...TCGGAACCTG 319  
||||| ||| :|||: ||| :|||  
1512 ProGlyProHisProLeuProAspProSerProAlaThrProLe 1528  
318 GCCATCGCTCCCGAGCC...ACGGAGCGCGCCACCGCTCACCAGC 272  
||| ||||| ||| :|||: ||| :|||: ||| :|||  
1528 uProAlaProProAlaValAlaAlaProProThrLeuAsps 1545  
271 ATCCGGCAACAGCC.....GCCGGATCACC 246  
: |||: ||| :|||: ||| :|||  
1545 erThrAlaSerSerLeuThrSerTyrAspSerGluValAlaThrLeuThr 1561  
245 GAGGGGCAACCGCTTTCGATCAGTCAGTCAGTCAGTCAGTCAGTCAG 196  
: |||: ||| :|||: ||| :|||  
1562 GlnGlyAlaPro.....AlaAlaPr 1568  
195 CAACGACCCCTCGCCAGCT.....AGGACTCCGCGCGCAGCAGC 152  
: |||: ||| :|||: ||| :|||  
1568 oGlyAspProAlaProGlyProProAlaProAlaProAlaProp 1585  
151 CCGGCGCGCGCTGGGCTGATCCACGAGCAGCGGATGTTGCACAGC 102  
||||| ||| ||||| ||| :|||: ||| :|||  
1585 roAlaProGlnProGlyProAspProProGly.....ThrAspSer 1599  
101 GACTG.....GTCCGAGCAGG.....CCATCTGCGC 73  
|||: ||| ||||| |||: ||| :||| :|||  
1600 GlyIleGluGluValAspSerArgSerSerSerAspHisProLeuGluTh 1616  
72 GGCTTCCTCGCTGGGTG 51  
: |||||: ||| :|||  
1616 rIleSerSerAlaSerThrLeu 1623

seq\_name: sp\_rodent:Q9WU13

seq\_documentation\_block:

ID Q9WU13 PRELIMINARY; PRT; 2158 AA.  
AC Q9WU13;  
DT 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE SYNAPTIC SAPAP-INTERACTING PROTEIN SYNAMON.  
OS Rattus norvegicus (Rat)  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=101116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99419021; PubMed=10488079;  
RA Yao I., Hata Y., Hirao K., Deguchi M., Ide N., Takeuchi M., Takai Y.;  
RT "Synamon, a novel neuronal protein interacting with synapse-associated  
J. Biol. Chem. 274:27463-27466(1999).  
RL EMBL: AF102855; AAD04569.2; -.  
DR HSSP: P00519; IABL  
DR InterPro: IPR002110; ANK.  
DR InterPro: IPR001478; PDZ.  
DR InterPro: IPR002965; P-rich\_extensn.  
DR InterPro: IPR001660; SAM.  
DR InterPro: IPR001452; SH3.  
DR Pfam: PF00023; ank; 6.  
DR Pfam: PF00595; PDZ; 1.  
DR Pfam: PF00536; SAM; 1.  
DR Pfam: PF00018; SH3; 1.  
DR PRINTS; PR01217; PRICHEXTENS.  
DR SMART; SM00248; ANK; 3.  
DR SMART; SM00228; PDZ; 1.  
DR SMART; SM00454; SAM; 1.  
DR SMART; SM00326; SH3; 1.  
DR PROSITE; PS50088; ANK\_REPEAT; 3.  
DR PROSITE; PS50297; ANK\_REPEAT\_REGION; 1.  
DR PROSITE; PS50106; PDZ; 1.  
DR PROSITE; PS50002; SH3; 1.  
KW ANK repeat; Repeat.  
SQ SEQUENCE 2158 AA; 225519 MW; 2AB6D53B5F1A4407 CRC64;

alignment\_scores:

Quality: 127.50 Length: 124  
Ratio: 1.903 Gaps: 9  
Percent Similarity: 54.032 Percent Identity: 36.290

alignment\_block:

US-09-462-480-3/rev x Q9WU13

Align seg 1/1 to: Q9WU13 from: 1 to: 2158

353 CCCGGG.....CTGGTGGAGCCGCCGAT...TCGGAACCTG 319  
||||| ||| :|||: ||| :|||  
1583 ProGlyProHisProLeuProAspProSerProAlaThrProLe 1599  
318 GCCATCGCTCCCGAGCC...ACGGAGCGCGCCACCGCTCACCAGC 272  
||| ||||| ||| :|||: ||| :|||: ||| :|||  
1599 uProAlaProProAlaValAlaAlaProProThrLeuAsps 1616  
271 ATCCGGCAACAGCC.....GCCGGATCACC 246  
: |||: ||| :|||: ||| :|||  
1616 erThrAlaSerSerLeuThrSerTyrAspSerGluValAlaThrLeuThr 1632  
245 GAGGGGCAACCGCTTTCGATCAGTCAGTCAGTCAGTCAGTCAGTCAG 196  
: |||: ||| :|||: ||| :|||  
1633 GlnGlyAlaPro.....AlaAlaPr 1639  
195 CAACGACCCCTCGCCAGCT.....AGGACTCCGCGCGCAGCAGC 152  
: |||: ||| :|||: ||| :|||  
1639 oGlyAspProAlaProGlyProProAlaProAlaProAlaProp 1656  
151 CCGGCGCGCGCTGGGCTGATCCACGAGCAGCGGATGTTGCACAGC 102  
||||| ||| ||||| ||| :|||: ||| :|||  
1656 roAlaProGlnProGlyProAspProProGly.....ThrAspSer 1670  
101 GACTG.....GTCCGAGCAGG.....CCATCTGCGC 73  
|||: ||| ||||| |||: ||| :||| :|||  
1671 GlyIleGluGluValAspSerArgSerSerSerAspHisProLeuGluTh 1687  
72 GGCTTCCTCGCTGGGTG 51  
: |||||: ||| :|||

1687 rIleSerSerAlaSerThrLeu 1694

seq\_name: sp\_rodent:Q9WV48

seq\_documentation\_block:

ID Q9WV48 PRELIMINARY; PRT; 2167 AA.  
AC Q9WV48;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE SPANK-1  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20549637; PubMed=10958799;  
RA Tobaben S., Sudhof T.C., Stahl B.;  
RT "The G protein-coupled Receptor CL1 Interacts Directly with Proteins  
of the Shank Family.";  
RL J. Biol. Chem. 275:36204-36210(2000).  
DR EMBL; AF159046; AD42975.1; -;  
DR HSSP; P00519; IABL.  
DR InterPro; IPR002110; ANK.  
DR InterPro; IPR001478; PDZ.  
DR InterPro; IPR002965; P\_rich\_extensn.  
DR InterPro; IPR001660; SAM.  
DR InterPro; IPR001452; SH3.  
DR Pfam; PF00023; ank; 6.  
DR Pfam; PF00595; PDZ; 1.  
DR Pfam; PF00536; SAM; 1.  
DR Pfam; PF00018; SH3; 1.  
DR PRINTS; PR01217; PRICHEXTENS.  
DR SMART; SM00248; ANK; 3.  
DR SMART; SM00228; PDZ; 1.  
DR SMART; SM00454; SAM; 1.  
DR SMART; SM00326; SH3; 1.  
DR PROSITE; PS00088; ANK\_REPEAT; 3.  
DR PROSITE; PS0297; ANK\_REPEAT\_REGION; 1.  
DR PROSITE; PS0106; PDZ; 1.  
DR PROSITE; PS0002; SH3; 1.  
KW ANK repeat; Repeat.  
SQ SEQUENCE 2167 AA; 226333 MW; 3F478B5A7B18BA86 CRC64;

alignment\_scores:  
Quality: 127.50 Length: 124  
Ratio: 1.903 Gaps: 9  
Percent Similarity: 54.032 Percent Identity: 36.290

alignment\_block:

US-09-462-480-3/rev x Q9WV48 ..

Align seg 1/1 to: Q9WV48 from: 1 to: 2167

353 CCGGGG.....CTGGTGGAGCGCGGAT...TGCGAACCGCTG 319  
|||||  
1592 ProGlyProProHisProLeuProAspProSerProAlaThrProLe 1608  
318 GCCCATCGTCCCGACCC...ACGGAGCGGCCACCGCGTCACCGACG 272  
|||||  
1608 uProAlaAlaProProAlaValAlaAlaAlaProProThrLeuAsp 1625  
271 ATCCGGCAACAGCC.....GCCGGCATCACC 246  
:::|||||  
1625 erThrAlaSerLeuThrSerTyrAspSerGluValAlaThrLeuThr 1641  
245 GAGGGGCAACCGGCTTTCAGTCAGCTAGACATCAGCGCGGTGGGGT 196  
:::|||||  
1642 GlnGlyAlaPro.....AlaAlaPr 1648  
195 CAACGACCCACCTCGCCAGGT.....AGCGACTCCGCGCGCAGCAGGC 152

1648 oGlyAspProProAlaProGlyProProAlaProAlaProAlaProp 1665  
:::|||||  
151 CCGCGCGCGCTGGGGCTGATCCACCGACCGGATGGTTGCAGACG 102  
|||||  
1665 roAlaProGlnProGlyProAspProProGly.....ThrAspSer 1679  
101 GGACTG.....GTGCGGACGAGG.....CCCATCTCGCG 73  
|||||  
1680 GlyIleGluGluValAspSerArgSerSerAspHisProLeuGluTh 1696  
72 GGCTTCTCTCGGCTGGGTG 51  
:|||||  
1696 rIleSerSerAlaSerThrLeu 1703

seq\_name: sp\_rodent:Q9ESZ9

seq\_documentation\_block:

ID Q9ESZ9 PRELIMINARY; PRT; 810 AA.  
AC Q9ESZ9;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
DE ELN (FRAGMENT).  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=129/SV;  
RA Green E.D.;  
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF289665; AAF99336.1; -;  
FT NON\_TER 1  
SQ SEQUENCE 810 AA; 67204 MW; 1ED61ED5AD4F40A2 CRC64;

alignment\_scores:  
Quality: 127.00 Length: 141  
Ratio: 1.649 Gaps: 8  
Percent Similarity: 54.610 Percent Identity: 34.752

alignment\_block:

US-09-462-480-3 x Q9ESZ9 ..

Align seg 1/1 to: Q9ESZ9 from: 1 to: 810

31 GTGGCGCGCACCGGC.....GGCGCAACCC 56  
:::|||||  
3 IleGlyGlyLeuGlyGlyGlyGlyAlaLeuGlyProGlyGlyLysPr 19  
57 AGCCGACGAGGAGCGCGCAGATGGCCCTGCTCGGCACCATCGCTGT 106  
|  
19 oProLysProGlyAlaGlyLeuGlyThrPheGlyAlaGlyProGlyG 36  
107 CGAACCATCGCTGGCTGGTGATCAGCCCGCGCGCGCGCGCTG 156  
|  
36 ly.....LeuGlyGlyAlaGlyProGlyAlaGlyLeuGlyAla 48  
157 CTGCGCGCGGAGTCGCTACCTGGCGCAGGTGGGTGTTGACCCG..... 200  
::: |||  
49 PheProAlaGlyThrPheProGlyAlaGlyAlaLeuVal..ProGlyGly 64  
201 .....CACGGCGCTGATGCTCAGCTGATCGAAAGCCGGTGGCC 241  
|||||  
65 AlaAlaGlyAlaAlaAlaAlaTyrLysAlaAlaAlaLysAlaGlyAlaGl 81  
242 CCTCGGTGATCGCGCGCTGTTCCCGGA.....TCGTCGGTG 279  
|||||  
81 yLeuGlyGlyValGlyGly.ValProGlyGlyValGlyValGlyVal 97  
280 ACGGTGGCGCGCTCGCGTGGGT...CCGGAGCGATGGCCAGGGTTC 326



OM of: US-09-462-480-2 to: A\_Geneseq\_032802:\* out\_format : pfs  
Date: Jul 22, 2002 1:22 AM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+np.model -DEV=xlh  
-O=/cgn2.1/USPTO\_spool/US09462480/runat\_18072002\_164418\_19413/app\_query.fasta\_1.2850  
-DB=A\_Geneseq\_032802 -QFMT=fastan -SUFFIX=rag -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000  
-CGAPOP=4.500 -CGAPEXT=0.050 -YGAPOP=10.000 -YGAPEXT=0.500  
-YGAPOP=6.000 -YGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62  
-TRANS=human40 cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORM=ext -HEAPSIZB=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09462480 @CGN1\_1.57 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIME=OUT=30 -NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-462-480-2  
Query length: 524  
Database: A\_Geneseq\_032802:\*  
Database sequences: 747574  
Database length: 111073796  
Search time (sec): 134.850000

score\_list:

Sequence	Strd Orig	ZScore	Escore	Len	Documentation
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:AAW32452 +	713.00	1059.80	1.9e-51		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:AAW32384 +	713.00	1059.80	1.9e-51		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW81705 +	713.00	1059.80	1.9e-51		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW64338 +	713.00	1059.80	1.9e-51		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW39135 +	713.00	1059.80	1.9e-51		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW38952 +	713.00	1059.80	1.9e-51		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:AAW19844 +	713.00	1059.80	1.9e-51		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW72929 +	713.00	1059.72	1.9e-51		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW21946 +	713.00	1059.72	1.9e-51		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW72948 +	373.00	559.42	3.7e-23		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW21945 +	373.00	559.42	3.7e-23		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:AAE01594 +	147.50	214.74	0.0003		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:AAW54248 +	145.50	215.70	0.0004		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:AAW79841 +	141.50	194.83	0.0010		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW24091 +	140.00	199.36	0.0012		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:ABG21932 +	136.50	193.02	0.0023		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:AAW31855 +	136.50	190.49	0.0024		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:AAW31852 +	135.50	200.47	0.0025		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:ABG13470 +	135.50	198.27	0.0026		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:ABG14738 +	133.00	185.05	0.0047		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW20852 +	132.50	196.20	0.0044		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW72204 +	132.50	172.16	0.0062		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:AAW39141 +	131.00	182.93	0.0069		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:ABG14734 +	128.50	175.80	0.0117		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:ABG21919 +	127.50	184.50	0.0122		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:ABG30150 +	127.50	182.36	0.0126		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:ABG31516 +	127.50	168.19	0.0154		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:AAW12000 +	127.50	168.18	0.0154		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:AAW18664 +	126.50	181.50	0.0151		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:AAW60620 +	125.50	165.76	0.0224		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:ABG03626 +	125.00	183.38	0.0190		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:ABW70063 +	124.50	177.24	0.0226		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:ABW1016 +	124.00	174.78	0.0255		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW29514 +	123.50	173.21	0.0284		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:AAW44463 +	123.50	173.21	0.0284		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:AAE13804 +	123.50	173.21	0.0284		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:ABG18528 +	123.00	179.52	0.0283		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:ABG14000 +	123.00	172.42	0.0313		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:ABW65293 +	122.50	165.62	0.0375		

/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:ABW61602 - 122.00 171.75 0.0  
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:ABG04558 - 121.00 176.14 0.0  
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:ABG19764 - 121.00 174.96 0.0  
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW68408 - 121.00 172.86 0.0  
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:AAW28150 - 121.00 172.14 0.0

seq\_name: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:AAW32452

seq\_documentation\_block:

ID AAW32452 standard; Protein; 368 AA.

XX AAW32452;

XX 09-JAN-1998 (first entry)

DE Mycobacterium tuberculosis antigen Tb37-FL.

KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;

KW skin testing; M.tuberculosis.

XX Mycobacterium tuberculosis.

XX Key Location/Qualifiers

FT Misc-difference 10

FT /note= "Any amino acid"

PN W09709428-A2.

XX 13-MAR-1997.

XX 30-AUG-1996; 96WO-US14674.

PR 12-JUL-1996; 96US-0680574.

PR 01-SEP-1995; 95US-0523436.

PR 22-SEP-1995; 95US-0533634.

PR 02-MAR-1996; 96US-0620874.

XX 05-JUN-1996; 96US-0659683.

PA (CORI-) CORIXA CORP.

PI Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;

PI Twardzik DR, Vedvick TH;

XX WPI; 1997-192903/17.

PT New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are  
useful in vaccines for prevention or treatment of tuberculosis, also  
for diagnosis

PS Example 3; Page 146-147; 168pp; English.

CC A new immunogenic polypeptide has been developed comprising an  
immunogenic part of a soluble Mycobacterium tuberculosis antigen (or  
its variant differing only in conservative substitutions and/or  
modifications). The present sequence represents a M.tuberculosis  
antigen, Tb37-FL, the immunogenic protein, and fusion proteins  
containing one or more of the proteins or one of the proteins plus  
ESAT-6, are useful in vaccines, preferably when formulated with a  
non-specific adjuvant, to induce an immune response against  
M.tuberculosis (for treatment or prevention).

XX Sequence 368 AA;

alignment\_scores: Quality: 713.00 Length: 143  
Ratio: 5.057 Gaps: 0  
Percent Similarity: 98.601 Percent Identity: 96.503

alignment\_block:

US-09-462-480-2 x AAW32452 ..

Align seg 1/1 to: AAW32452 from: 1 to: 368

```
1 CTGCAGCAGGTGACCTGCTGTTTACAGCAGGTGGCGGCACCGCGCGG 50
|||||
236 LeuGlnGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGlyG1 242
51 CAACCCAGCCGACGAGGAGCGGCACATGGCCCTGCTCGGCACCATC 100
|||||
242 yAsnProAlaAspGluGluAlaAlaGlnMetGlyLeuLeuGlyThrSerP 259
101 CGCTGTCGAACCATCGCTGGTGTGATCAGGCCCGCCAGCGCGCGCG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlySerGlyProSerAlaGlyAla 275
151 GGCCTGTCGCGCGGAGTACCTGCTGCGCAGGAGTGGTGTGACCCG 200
|||||
276 GlyLeuLeuArgAlaGlySerLeuProGlyAlaGlyGlySerLeuThrAr 292
201 CACGCCGCTGATGCTCAGCTCATCGAAAGCCGTTGCCCTCCGTTGA 250
|||||
292 gThrProLeuMetSerGlnLeuLeuGlyProValAlaProSerValM 309
251 TGCCGCGCGCTGTTCCGCGATCGTGGTGGTGGCGCGCTCGCGTG 300
|||||
309 etProAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
301 GGTCCGCGGAGCGTGGCGAGGTTCGCAATCCGCGCGCTCCACGCCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
351 GGGTCTGCTGCGCGCGGCGCGCTGCGCAGGAGCGTGAAGAAGACGACG 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAspG 359
401 AGGACGACTGGGACGAGGAGGACGACTGG 429
|||||
359 luAspAspTrpAspGluAspAspTrp 368
```

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:AAW32384

seq\_documentation\_block:

ID AAW32384 standard; Protein; 368 AA.

AC AAW32384;

DT 13-JAN-1998 (first entry)

DE Mycobacterium tuberculosis antigen Tb37-FL.

KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;

KW skin testing; M.tuberculosis.

OS Mycobacterium tuberculosis.

FH Key Location/Qualifiers

FT Misc-difference 10 /note= "Any amino acid"

FT WO9709429-A2.

PN 13-MAR-1997.

XX 30-AUG-1996; 96WO-0514675.

XX 12-JUL-1996; 96US-0680573.

PR 01-SEP-1995; 95US-0523435.

PR 22-SEP-1995; 95US-0532136.

PR 22-MAR-1996; 96US-0620280.

PR 05-JUN-1996; 96US-0658800.

XX (CORI-) CORIXA CORP.

XX Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;

PI Twardzik DR, Vedvick TH;

PI

```
XX WPI; 1997-192904/17.
XX DR
XX PT
XX - useful for diagnosis of M. tuberculosis infection
XX PS
XX Example 3; Page 159-161; 190pp; English.
XX CC
XX A new immunogenic polypeptide has been developed comprising an
XX immunogenic part of a soluble Mycobacterium tuberculosis antigen (or
XX its variant differing only in conservative substitutions and/or
XX modifications). The present sequence represents a M.tuberculosis
XX antigen, Tb37-FL. The immunogenic polypeptide can be used to diagnose
XX M.tuberculosis infection by forming complexes with specific
XX antibodies in the sample. Fragments of DNA encoding the immunogenic
XX polypeptide can be used as diagnostic primers or probes and agents
XX that bind to the antigen, especially monoclonal antibodies or
XX equivalent polyclonal antibodies, are also used for diagnosis.
XX SQ Sequence 368 AA;
```

alignment\_scores:

Quality: 713.00 Length: 143

Ratio: 5.057 Gaps: 0

Percent Similarity: 98.601 Percent Identity: 96.503

alignment\_block:

US-09-462-480-2 x AAW32384 ..

Align seg 1/1 to: AAW32384 from: 1 to: 368

```
1 CTGCAGCAGGTGACCTGCTGTTTACAGCAGGTGGCGGCACCGCGCGG 50
|||||
236 LeuGlnGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGlyG1 242
51 CAACCCAGCCGACGAGGAGCGGCACATGGCCCTGCTCGGCACCATC 100
|||||
242 yAsnProAlaAspGluGluAlaAlaGlnMetGlyLeuLeuGlyThrSerP 259
101 CGCTGTCGAACCATCGCTGGTGTGATCAGGCCCGCCAGCGCGCGCG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlySerGlyProSerAlaGlyAla 275
151 GGCCTGCTGCGCGGAGTACCTGCTGCGCAGGAGTGGTGTGACCCG 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292
201 CACGCCGCTGATGCTCAGCTCATCGAAAGCCGTTGCCCTCCGTTGA 250
|||||
292 gThrProLeuMetSerGlnLeuLeuGlyProValAlaProSerValM 309
251 TGCCGCGCGCTGTTCCGCGATCGTGGTGGTGGCGCGCTCGCGTG 300
|||||
309 etProAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
301 GGTCCGCGGAGCGTGGCGAGGTTCGCAATCCGCGCGCTCCACGCCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
351 GGGTCTGCTGCGCGCGGCGCGCTGCGCAGGAGCGTGAAGAAGACGACG 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAspG 359
401 AGGACGACTGGGACGAGGAGGACGACTGG 429
|||||
359 luAspAspTrpAspGluAspAspTrp 368
```

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW81705

seq\_documentation\_block:

ID AAW81705 standard; Protein; 368 AA.

XX

```

AC AAW81705;
XX
XX 27-JAN-1999 (first entry)
XX
DE M. tuberculosis immunogenic polypeptide Tb37-FL.
XX
XX Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
XX KW vaccine; pharmaceutical; infection; diagnosis.
XX
XX Mycobacterium tuberculosis.
XX
XX Key Location/Qualifiers
FH Misc-difference 10
FT /label= unknown
XX
XX WO9816646-A2.
XX
XX 23-APR-1998.
XX
XX 07-OCT-1997; 97WO-US18293.
XX
XX 13-MAR-1997; 97US-0818112.
XX PR 11-OCT-1996; 96US-0730510.
XX
XX (CORI-) CORIXA CORP.
XX
XX Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
XX PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX
XX WPI; 1998-261042/23.
XX
XX Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used
XX to develop products for the detection of M. tuberculosis infection
XX and for diagnosis, treatment and prevention of tuberculosis
XX
XX Example 3B; Page 137-138; 230pp; English.
XX
XX This sequence represents an immunogenic portion of a soluble
XX Mycobacterium tuberculosis (MT) antigen which can be used in a method
XX for inducing protective immunity against tuberculosis (TB). This
XX sequence can be formulated into vaccines and/or pharmaceutical
XX compositions for immunising against M. tuberculosis infection or may
XX be used for the diagnosis of tuberculosis.
XX
XX Sequence 368 AA;

alignment_scores:
Quality: 713.00 Length: 143
Ratio: 5.057 Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503

alignment_block:
US-09-462-480-2 x AAW81705 ..
Align seg 1/1 to: AAW81705 from: 1 to: 368
1 CTGCAGCAGGTGACGTGCTGTTCTTCAGCCAGGTGGCGGCACCGCGCGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGlyGly 242
51 CAACCCAGCCGACGAGGAGCCGCGCAGATGGGCGCTGCTCGGCACCAAGTC 100
|||||
242 YasnProAlaAspGluGluAlaAlaGlnMetGlyLeuLeuGlyThrSerP 259
101 CGCTGTGCAACATCCGCTGCTGGTGGATCAGCCCGCCAGCGCGGCGCG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275
151 GGCCTGTGCGCGCGGAGTCCCTACCTGGCGCAGGTGGTGGTGCACCCG 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292

```

```

201 CACGCCGCTGATGTCTCAGCTGATCGAAAGCCGTTGCCCCCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309
251 TGGCGGCGGCTGTTCGCGGATCGTGGTGACGGGTGGCGCGCTCCGGTG 300
|||||
309 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
301 GGTCGCGGAGCGATGGCGGTTTCGCAATCCGCGGCTCCACCGACGCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
351 GGGTCTGTGTCGCGCGCCGACCGCTCGCGGAGGACGCTGAAGAAGACGAC 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspG 359
401 AGGACGACTGGGACGAAAGAGGACGACTGG 429
|||||
359 luAspAspTrpAspGluGluAspAspTrp 368

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1998.DAT:AAW64338
seq_documentation_block:
ID AAW64338 standard; Protein; 368 AA.
XX
XX AAW64338;
XX
XX 09-NOV-1998 (first entry)
XX
XX Mycobacterium tuberculosis antigen Tb37-FL.
XX
XX Tuberculosis; infection; diagnosis; antigen; Tb37-FL.
XX
XX Mycobacterium tuberculosis strain H37Rv.
XX
XX Key Location/Qualifiers
FH Misc-difference 10 /note= "unidentified"
XX
XX WO9816645-A2.
XX
XX 23-APR-1998.
XX
XX 07-OCT-1997; 97WO-US18214.
XX
XX 13-MAR-1997; 97US-0818111.
XX PR 11-OCT-1996; 96US-0729622.
XX
XX (CORI-) CORIXA CORP.
XX
XX Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
XX PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX
XX WPI; 1998-251292/22.
XX
XX New isolated Mycobacterium tuberculosis polypeptides and DNA - used
XX to develop products for the detection of M. tuberculosis infection
XX and diagnosis of tuberculosis
XX
XX Example 3; Page 143-144; 250pp; English.
XX
XX This polypeptide comprises Mycobacterium tuberculosis antigen
XX Tb37-FL. It is encoded by genomic DNA isolated from a M.
XX tuberculosis strain H37Rv genomic library using a probe from
XX clone Tb38-1 (see AAW44384). The invention relates to compositions
XX and methods for diagnosing tuberculosis. It provides polypeptides
XX (see AAW64291-W64379) comprising an antigenic portion of a soluble
XX M. tuberculosis antigen, or an immunogenic portion of an M.
XX tuberculosis antigen, as well as DNA sequences encoding such
XX polypeptides, recombinant expression vectors and transformed or
XX kits for detecting M. tuberculosis infection in a patient using
XX these polypeptides, antibodies or oligonucleotide probes and

```

CC primers, for the diagnosis of tuberculosis.

XX  
SQ Sequence 368 AA;

alignment\_scores:

Quality: 713.00 Length: 143  
Ratio: 5.057 Gaps: 0  
Percent Similarity: 98.601 Percent Identity: 96.503

alignment\_block:

US-09-462-480-2 x AAW64338 ..

Align seg 1/1 to: AAW64338 from: 1 to: 368

```
1 CTGCAGCAGGTGACGTCGCTGTTTCAGCCAGGTGGCGGCACCGCGGGCGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGlyG 242
51 CAACCCAGCCGACGAGGAGCGCGCAGATGGCTGCTCGGCACCAATC 100
|||||
242 yAsnProAlaAspGluGluAlaGlnMetGlyLeuLeuGlyThrSerP 259
101 CGCTGTCAACCATCCGCTGCTGCTGATCAGCCCGCCAGCGGGCGCG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlySerGlyProSerAlaGlyAla 275
151 GGCTGTGTCGCGCGGAGTCTGCTACCTACCTGGCGGAGTGGTGGTGC 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlySerLeuThrAr 292
201 CACGCCGCTGATGCTCAGCTGATCGAAGCCGGTGGCCCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309
251 TGCCGGCGGCTGTTCGCGGATCGCTGCTGAGCGGTGGCGCGCTCGGTG 300
|||||
309 etProAlaAlaAlaGlySerAlaThrGlyGlyAlaAlaProVal 325
301 GGTCGGGAGCGAGTGGCCAGGTTCCCAATCCGCGGCTCCACCGCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlySerThrArgrP 342
351 GGGTCTGTGTCGCGCGCACCGCTCGCGCAGGAGCGGTGAAGAACGACG 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspG 359
401 AGGACGACTGGGACGAGGACGACTGG 429
359 luAspAspTrpAspGluGluAspAspTrp 368
```

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAV39135.

seq\_documentation\_block:

ID AAV39135 standard; Protein; 368 AA.

XX  
AC AAV39135;

XX  
DT 05-NOV-1999 (first entry)

DE M. tuberculosis antigen Tb37-FL amino acid sequence.

XX  
KW Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;  
KW immunotherapy; diagnosis; immunisation; vaccine; infection;  
KW immune response; skin test.

XX  
OS Mycobacterium tuberculosis.

XX  
PN W09942076-A2.

XX  
PD 26-AUG-1999.

XX  
PF 17-FEB-1999; 99WO-US03268.

XX  
PR 05-MAY-1998; 98US-0072967.  
PR 18-FEB-1998; 98US-0025197.

XX  
PA (CORI-) CORIXA CORP.

XX  
PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;  
PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedwick TS;

XX  
DR WPI; 1999-527409/44.

XX  
PT New antigens from Mycobacterium tuberculosis useful in diagnostic  
PT skin tests and protective or therapeutic vaccines or compositions

XX  
PS Example 3; Page 132-133; 299pp; English.

XX  
CC The present invention describes polypeptides comprising an immunogenic  
CC part of a Mycobacterium tuberculosis antigen (Ag). Also described  
CC are vaccines and fusion protein containing M. tuberculosis Ag's.  
CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and  
CC other polypeptides fragments, can be used in pharmaceutical compositions  
CC or vaccines to generate a protective or therapeutic immune response to  
CC M. tuberculosis and as reagents in skin tests for diagnosis of  
CC tuberculosis. Ag can induce proliferation of, or cytokine secretion  
CC by, T, B or natural killer cells and/or macrophages in  
CC tuberculosis-immune subjects. AA19249 to AA219460 and AAV39083 to  
CC AAV39225 are used in the exemplification of the present invention.

XX  
SQ Sequence 368 AA;

alignment\_scores:

Quality: 713.00 Length: 143  
Ratio: 5.057 Gaps: 0  
Percent Similarity: 98.601 Percent Identity: 96.503

alignment\_block:

US-09-462-480-2 x AAV39135 ..

Align seg 1/1 to: AAV39135 from: 1 to: 368

```
1 CTGCAGCAGGTGACGTCGCTGTTTCAGCCAGGTGGCGGCACCGCGGGCGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGlyG 242
51 CAACCCAGCCGACGAGGAGCGCGCAGATGGCTGCTCGGCACCAATC 100
|||||
242 yAsnProAlaAspGluGluAlaGlnMetGlyLeuLeuGlyThrSerP 259
101 CGCTGTCAACCATCCGCTGCTGCTGATCAGCCCGCCAGCGGGCGCG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlySerGlyProSerAlaGlyAla 275
151 GGCTGTGTCGCGCGGAGTGGCCAGGTTCCCAATCCGCGGCTCCACCGCC 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlySerLeuThrAr 292
201 CACGCCGCTGATGCTCAGCTGATCGAAGCCGGTGGCCCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309
251 TGCCGGCGGCTGTTCGCGGATCGCTGCTGAGCGGTGGCGCGCTCGGTG 300
|||||
309 etProAlaAlaAlaGlySerAlaThrGlyGlyAlaAlaProVal 325
301 GGTCGGGAGCGAGTGGCCAGGTTCCCAATCCGCGGCTCCACCGCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlySerThrArgrP 342
351 GGGTCTGTGTCGCGCGCACCGCTCGCGCAGGAGCGGTGAAGAACGACG 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspG 359
```



401 AGGACGACTGGGACGAAGAGGACGACTGG 429  
|||||  
359 luAspAspTTPaspGluGluAspAspTTP 368

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AA199992

seq\_documentation\_block:

ID AAY38992 standard; Protein; 368 AA.  
AC AAY38992;  
XX  
DT 05-NOV-1999 (first entry)  
XX  
DE M. tuberculosis recombinant antigen protein Tb37-FL.  
XX  
KW Antigen; diagnosis; detection; infection; antibody; immunisation;  
KW vaccine; immunity.  
XX  
OS Mycobacterium tuberculosis.  
XX  
PN WO99421118-A2.  
PN  
PD 26-AUG-1999.  
XX  
PF 17-FEB-1999; 99WO-US03265.  
XX  
PR 05-MAY-1998; 98US-0072596.  
PR 18-FEB-1998; 98US-0024753.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;  
PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;  
XX  
DR WPI; 1999-527416/44.  
XX  
PT New polypeptide comprising antigenic portions of M. tuberculosis  
XX  
PS Example 3; Page 177-179; 323pp; English.  
XX  
CC This invention describes novel recombinant antigens and their encoding  
CC nucleic acids derived from Mycobacterium tuberculosis. The novel  
CC polypeptides are useful for detecting M. tuberculosis infection in a  
CC biological sample by detecting antibodies which bind with the  
CC polypeptides, and are useful as vaccines for immunizing against  
CC M. tuberculosis infection. The new detection methods are needed as  
CC current vaccination strategies do not provide 100% immunity.  
XX  
SQ Sequence 368 AA;

alignment\_scores:  
Quality: 713.00 Length: 143  
Ratio: 5.057 Gaps: 0  
Percent Similarity: 98.601 Percent Identity: 96.503  
alignment\_block:  
US-09-462-480-2 x AAY38992 ..  
Align seg 1/1 to: AAY38992 from: 1 to: 368  
1 CTGCAGCAGGTGACGTCTGTTTCAGCCAGGTGGCGGACCGCGCGG 50  
|||||  
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGlyG 242  
|||||  
51 CAACCCAGCCGACGAGGAAGCCGCGAGATGGGCTGCTCGGCACCAATC 100  
|||||  
242 YasnProAlaAspGluGluAlaAlaGlnMetGlyLeuLeuGlyThrSerP 259  
|||||  
101 CGCTGTGCAACATCCGCTGCTGGTGTGATCAGGCCCGCGCGGCGG 150  
|||||  
259 roLeuSerAsnHisProLeuAlaGlySerGlyProSerAlaGlyAla 275  
|||||

151 GGCTCTGCTGGCGCGGAGTGCTACCTGCGGAGGTGGTGGTGGTGGTGG 200  
|||||  
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292  
|||||  
201 CAGCCGCTGATGCTCTCAGTGATCGAAAGCCGGTTGCCCTCGGTGA 250  
|||||  
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309  
|||||  
251 TGCGCGCGGCTGTGGCGGATCGTCGTCGAGGGTGGCGGCTGCCGGTG 300  
|||||  
309 etProAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325  
|||||  
301 GGTCCGGGAGCGATGGCCAGGGTTCGCAATCCGGCGGCTCCACCGCC 350  
|||||  
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342  
|||||  
351 GGTCTGCTGGCGCGGCGGCGGCTCGCGGAGGAGCGGTGAAGAACGACG 400  
|||||  
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAspG 359  
|||||  
401 AGGACGACTGGGACGAAGAGGACGACTGG 429  
|||||  
359 luAspAspTTPaspGluGluAspAspTTP 368

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:AA19844

seq\_documentation\_block:

ID AAB19844 standard; Protein; 368 AA.  
XX  
AC AAB19844;  
XX  
DT 05-MAR-2001 (first entry)  
XX  
DE Mycobacterium tuberculosis protein MTBN3.  
XX  
KW MTBN3; tuberculosis; BCG; vaccine; infection; diagnosis.  
XX  
OS Mycobacterium tuberculosis.  
XX  
PN WO2000066157-A1.  
XX  
PD 09-NOV-2000.  
XX  
PF 04-MAY-2000; 2000WO-US12257.  
XX  
PR 04-MAY-1999; 99US-0132505.  
XX  
PA (PUBL-) PUBLIC HEALTH RES INST NEW YORK.  
XX  
PI Gennaro ML;  
XX  
DR WPI; 2001-007153/01.  
DR N-PSDB; AAA89037.  
XX

Novel polypeptide encoded by open reading frames present in  
Mycobacterium tuberculosis genome and not by the BCG strain of M.  
bovis, useful as vaccine and for diagnosing tuberculosis infection  
XX  
PS Claim 11; Fig 1; 35pp; English.  
XX  
CC The present sequence is that of the Mycobacterium tuberculosis  
CC MTBN3 protein. This is 1 of 8 proteins, i.e. MTBN1-8 (see  
CC AAB19842-49), encoded by 8 open reading frames (see AAA89035-42)  
CC identified as being present in the genome of M. tuberculosis but  
CC absent from the genome of the BCG strain of Mycobacterium bovis.  
CC MTBN1-8 represent reagents that are useful in discriminating between  
CC M. tuberculosis and BCG and, in particular, for diagnostic methods  
CC which discriminate between exposure of a subject to M. tuberculosis  
CC and vaccination with BCG. The invention features these MTBN  
CC polypeptides, functional fragments of them, DNA encoding them,  
CC vectors, transformed cells, and diagnostic, therapeutic, and  
CC prophylactic (vaccine) methods, including genetic vaccination  
CC methods.



362 luAspAspTrpAspGluAspAspTrp 371

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1999.DAT:AAV21946

seq\_documentation\_block:

ID\_AAY21946 standard; Protein; 371 AA.

XX AC AAY21946;

XX DT 06-SEP-1999 (first entry)

XX DE Amino acid sequence of antigen RD1-ORF5.

XX KW Immunogenic; Mycobacterium tuberculosis; immune response; infection;  
KW tuberculosis; fusion polypeptide; T-cell epitope; ESAT-6; MPT59; TB;  
KW pharmacological; vaccination; M. africanum; M. bovis; CFP7A; CFP30A;  
KW CFP7B; CFP19; CFP27; CFP30A; RD1-ORF; CFP10A; CFP16; CFP19; CFP23;  
KW CFP25A; CFP30B; CFP7B.

XX OS Mycobacterium tuberculosis.

XX PN WO924577-Al.

XX PD 20-MAY-1999.

XX PF 08-OCT-1998; 98WO-DK00438.

XX PR 01-APR-1998; 98WO-DK00132.

XX PR 10-NOV-1997; 97DK-0001277.

XX PR 05-JAN-1998; 98US-0070488.

XX XX (STAT-) STATENS SERUM INST.

XX XX Andersen P, Skjot R;

XX DR WPI; 1999-347282/29.

XX DR N-PSDB; AAX81046.

XX PT New immunogenic fragment of Mycobacterium tuberculosis

XX PS Example 2; Page 219-220; 265pp; English.

XX CC The invention describes a substantially pure immunogenic polypeptide  
CC fragment (I) from Mycobacterium tuberculosis that is able to evoke a  
CC protective immune response against infections by mycobacteria belonging  
CC to the tuberculosis complex. The invention provides a (1) fusion  
CC polypeptide comprising at least one polypeptide fragment (I) and at least  
CC one fusion partner; (2) a fusion polypeptide fragment comprising a T-cell  
CC epitope from M. tuberculosis protein ESAT-6, or MPT59 and a second  
CC different amino acid sequence from M. tuberculosis, and/or including a  
CC sequence which protects the first amino acid sequence from in vivo  
CC degradation or post-translational processing; (3) a nucleic acid fragment  
CC that encodes the above polypeptides. The polypeptides and nucleic acid  
CC are useful as pharmaceuticals, for diagnosis of and as antigens for  
CC vaccination against TB caused by Mycobacterium tuberculosis, africanum or  
CC bovis. The polypeptides are also useful for diagnosing ongoing or  
CC previous sensitization in an animal with bacteria belonging to the  
CC tuberculosis complex. The invention also describes the use of CFP7A or  
CC CFP30A or a T-cell epitope of for the induction of a strong immune  
CC response in a mammal; use of CFP7B, CFP19 or MPT59-ESAT6 or a T-cell  
CC epitope of for diagnosis of TB in a mammal by performing a DTH type skin  
CC test; use of CFP27, CFP30A, RD1-ORF2, RD1-ORF3, RD1-ORF5, MPT59-ESAT6,  
CC ESAT6-MPT59, CFP10A, CFP16, CFP19, CFP23, CFP25A, CFP30B, CFP7B or a T-  
CC cell epitope of for the preparation of an immunological composition; and  
CC for the preparation of a subunit vaccine.

XX SQ Sequence 371 AA;

alignment\_scores:

Quality: 713.00 Length: 143  
Ratio: 5.057 Gaps: 0  
Percent Similarity: 98.601 Percent Identity: 96.503

alignment\_block:

US-09-462-480-2 x AAY21946 ..

Align seg 1/1 to: AAY21946 from: 1 to: 371

1 CTGCAGCAGGTGACGTCGTTGTTTCAGCCAGGTGGCGGCACCGCGCGG 50  
|||||  
229 LeuGlnGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGlyG 245  
51 CAACCCAGCCGACGAGGAGCCGCGGAGATGGGCGCTGCTGGCCACCA 100  
|||||  
245 YAsnProAlaAspGluGluAlaGlnMetGlyLeuLeuGlyThrSerP 262  
101 CGCTGTCTGAACCATCCGCTGGTGGATCAGCCGCCAGCGCGCGCG 150  
|||||  
262 rLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 278  
151 GGCTGCTGCGCGGAGTCTACCTGGCGCAGGTGGGTGGTTCACCCG 200  
|||||  
279 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThr 295  
201 CAGCCCGCTGATGTCAGCTGATCGAAAGCCGGTTGCCCGCTCGGTGA 250  
|||||  
295 gThrProLeuMetSerGlnLeuLeuLysProValAlaProSerValm 312  
251 TCGCGCGCGTGTTCGCGGATCGTCGTCGTCGTCGTCGTCGTCGTCG 300  
|||||  
312 etProAlaAlaAlaAlaGlySerAlaThrGlyGlyAlaAlaProVal 328  
301 GGTCCGCGGAGCGATGGCCAGGGTTTCGCAATCCGCGCGCTCCACCA 350  
||| |||||  
329 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArg 345  
351 GGGTCTGGTCCGCGCGCACCGCTCGCGCAGGAGCGGTGAAGAACAGC 400  
|||||  
345 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAsp 362  
401 AGGAGGACTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 429  
|||||  
362 luAspAspTrpAspGluAspAspTrp 371

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA1998.DAT:AAW72928

seq\_documentation\_block:

ID\_AAW72928 standard; Protein; 139 AA.

XX AC AAW72928;

XX DT 21-JAN-1999 (first entry)

XX DE Mycobacterium tuberculosis antigen RD1-ORF4.

XX KW Mycobacterium tuberculosis; antigen; vaccine; immunological;  
KW immunogen; infection.

XX OS Mycobacterium tuberculosis.

XX PN WO9844119-Al.

XX PD 08-OCT-1998.

XX PF 01-APR-1998; 98WO-DK00132.

XX PR 05-JAN-1998; 98US-0070488.

XX PR 02-APR-1997; 97DK-0000376.

XX PR 18-APR-1997; 97US-0044624.

XX PR 10-NOV-1997; 97DK-0001277.

XX XX (STAT-) STATENS SERUM INST.

XX XX Andersen P, Florio W, Nielsen R, Oettinger T, Rasmussen PB;

XX PI Rosenkrands I, Weidinger K;

```
XX DR WPI; 1998-542705/46.
XX DR N-PSDB; AAV63938.
XX PT New isolated mycobacteria polypeptides and nucleic acids - used for
XX PT developing products for the diagnosis of or vaccination against
XX PT mycobacterial infections, particularly tuberculosis
XX PS Claim 1; Page 198; 163pp; English.
XX CC The present sequence represents a Mycobacterium tuberculosis protein.
XX CC Products from the present invention, which describes protein fragments
XX CC and nucleic acid fragments derived from M.tuberculosis, can be used in
XX CC the detection of and prevention of mycobacterial infections. In
XX CC particular, the proteins and nucleic acids can be used for the diagnosis
XX CC of or vaccination against tuberculosis caused by M. tuberculosis,
XX CC M. africanum or M. bovis.
XX SQ Sequence 139 AA;

alignment_scores:
    Quality: 373.00      Length: 67
    Ratio: 5.567        Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.507

alignment_block:
US-09-462-480-2/rev x AAW72928 ..
Align seg 1/1 to: AAW72928 from: 1 to: 139
203 GTGCGGGTCAACGACCCACCTGCGCGAGTAGCGACTCCGCGCGCAGCAG 154
: : : : : : : : : : : : : : : : : : : : : : : : : : :
1 MetArgValAsnAspProProAlaProGlySerAspSerAlaArgSerAr 17
153 GCCGCGCGCGCGCTGGGGCTGATCCACCAGCCGCGGATGTCGACA 104
: : : : : : : : : : : : : : : : : : : : : : : : : : :
17 gProAlaProAlaLeuGlyProAspProProAlaSerGlyTrpPheAsps 34
103 CGCGACTGGTGCGCGAGCAGGCCCATCTCGCGGGCTCTCTCGTGGCTGGG 54
: : : : : : : : : : : : : : : : : : : : : : : : : : :
34 erGlyLeuValProSerArgProIleCysAlaAlaSerSerAlaGly 50
53 TTGCGCGCGCGGTGGCGCCCACTGGCTGAACACGACGTCACCTGCTG 4
: : : : : : : : : : : : : : : : : : : : : : : : : : :
51 LeuProProValProProThrTrpLeuAsnAsnAspValThrCysCy 67
3 C 3
67 s 67

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAV21945
seq_documentation_block:
ID: AAY21945 standard; Protein; 139 AA.
XX AC AAY21945;
XX DT 06-SEP-1999 (first entry)
XX DE Amino acid sequence of antigen RD1-ORF4.
XX KW Immunogenic; Mycobacterium tuberculosis; immune response; infection;
XX KW tuberculosis; fusion polypeptide; T-cell epitope; ESAT-6; MPT59; TB;
XX KW pharmaceutical; vaccination; M. africanum; M. bovis; CFP7A; CFP30A;
XX KW CFP7B; CFP19; CFP27; CFP30A; RD1-ORF; CFP10A; CFP16; CFP19; CFP23;
XX KW CFP25A; CFP30B; CFP7B.
XX OS Mycobacterium tuberculosis.
XX PN WO9924577-A1.
XX PN 20-MAY-1999.
XX PD ..
```

```
XX 08-OCT-1998; 98WO-DK00438.
XX 01-APR-1998; 98WO-DK00132.
XX 10-NOV-1997; 97DK-0001277.
XX 05-JAN-1998; 98US-0070488.
XX (STAT-) STATENS SERUM INST.
XX PI Andersen P, Skjot R;
XX WPI; 1999-347282/29.
XX DR N-PSDB; AAX81045.
XX PT New immunogenic fragment of Mycobacterium tuberculosis
XX PS Example 2; Page 216; 265pp; English.
XX CC The invention describes a substantially pure immunogenic polypeptide
XX CC fragment (1) from Mycobacterium tuberculosis that is able to evoke a
XX CC protective immune response against infections by mycobacteria belonging
XX CC to the tuberculosis complex. The invention provides a (1) fusion
XX CC polypeptide comprising at least one polypeptide fragment (1) and at least
XX CC one fusion partner; (2) a fusion polypeptide fragment comprising a T-cell
XX CC epitope from M. tuberculosis protein ESAT-6, or MPT59 and a second
XX CC different amino acid sequence from M. tuberculosis, and/or including a
XX CC sequence which protects the first amino acid sequence from in vivo
XX CC degradation or post-translational processing; (3) a nucleic acid fragment
XX CC that encodes the above polypeptides. The polypeptides and nucleic acid
XX CC are useful as pharmaceuticals, for diagnosis of and as antigens for
XX CC vaccination against TB caused by Mycobacterium tuberculosis, africanum or
XX CC bovis. The polypeptides are also useful for diagnosing ongoing or
XX CC previous sensitization in an animal with bacteria belonging to the
XX CC tuberculosis complex. The invention also describes the use of CFP7A or
XX CC CFP30A or a T-cell epitope of for the induction of a strong immune
XX CC response in a mammal; use of CFP7B, CFP19 or MPT59-ESAT6 or a T-cell
XX CC epitope of for diagnosis of TB in a mammal by performing a DTH type skin
XX CC test; use of CFP27, CFP30A, RD1-ORF2, RD1-ORF3, RD1-ORF5, MPT59-ESAT6,
XX CC ESAT6-MPT59, CFP10A, CFP16, CFP19, CFP23, CFP25A, CFP30B, CFP7B or a T-
XX CC cell epitope of for the preparation of an immunological composition; and
XX CC for the preparation of a subunit vaccine.
XX SQ Sequence 139 AA;

alignment_scores:
    Quality: 373.00      Length: 67
    Ratio: 5.567        Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.507

alignment_block:
US-09-462-480-2/rev x AAY21945 ..
Align seg 1/1 to: AAY21945 from: 1 to: 139
203 GTGCGGGTCAACGACCCACCTGCGCGAGTAGCGACTCCGCGCGCAGCAG 154
: : : : : : : : : : : : : : : : : : : : : : : : : : :
1 MetArgValAsnAspProProAlaProGlySerAspSerAlaArgSerAr 17
153 GCCGCGCGCGCGCTGGGGCTGATCCACCAGCCGCGGATGTCGACA 104
: : : : : : : : : : : : : : : : : : : : : : : : : : :
17 gProAlaProAlaLeuGlyProAspProProAlaSerGlyTrpPheAsps 34
103 CGCGACTGGTGCGCGAGCAGGCCCATCTCGCGGGCTCTCTCGTGGCTGGG 54
: : : : : : : : : : : : : : : : : : : : : : : : : : :
34 erGlyLeuValProSerArgProIleCysAlaAlaSerSerAlaGly 50
53 TTGCGCGCGCGGTGGCGCCCACTGGCTGAACACGACGTCACCTGCTG 4
: : : : : : : : : : : : : : : : : : : : : : : : : : :
51 LeuProProValProProThrTrpLeuAsnAsnAspValThrCysCy 67
3 C 3
67 s 67
```

67 s 67

```
seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:AAE01594
seq_documentation_block:
ID_ AAE01594 standard; Protein; 314 AA.
XX
AC_ AAE01594;
XX
DT_ 17-JUL-2001 (first entry)
XX
DE_ Human gene 12 encoded secreted protein HUSGW69, SEQ ID NO:144.
XX
KW_ Human; secreted protein; proliferative disorder; cancer; tumour;
KW_ foetal abnormality; developmental abnormality; haematopoietic disorder;
KW_ immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW_ inflammation; allergy; neurological disorder; Alzheimer's disease;
KW_ Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW_ skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW_ cardiovascular disorder; angiogenic disorder; kidney disorder;
KW_ gastrointestinal disorder; pregnancy-related disorder;
KW_ endocrine disorder; infection; wound healing; vulnerability;
KW_ cell culture; chemotaxis; food additive; chromosome 17;
KW_ binding partner identification; gene therapy.
XX
OS_ Homo sapiens.
XX
FH_ Key Location/Qualifiers
FT_ Peptide 1..26
FT_ /label= Signal_peptide
FT_ Protein 27..55
FT_ /label= Mature_human_secreted_protein
FT_ Misc-difference 129
FT_ /label= Unknown
FT_ /note= "Encoded by MCG"
XX
PN_ WO200134623-A1.
XX
PD_ 17-MAY-2001.
XX
PF_ 01-NOV-2000; 2000WO-US30037.
XX
PR_ 05-NOV-1999; 98US-0163577.
XX
PR_ 30-JUN-2000; 2000US-0215137.
XX
PA_ (HUMA-) HUMAN GENOME SCI INC.
XX
PI_ Ruben SM, Komatsoulis GA, Moore PA;
XX
DR_ WPI; 2001-316490/33.
XX
DR_ N-PSDB; AAD05437.
XX
PT_ Nucleic acids encoding 29 human secreted polypeptides, useful for
PT_ preventing, diagnosing and/or treating e.g. cancers, Parkinson's
PT_ disease and diabetic retinopathy -
XX
PS_ Claim 11; Page 496-497; 535pp; English.
XX
CC_ AAD05389-AAD05473 represent cDNAs corresponding to 29 human secreted
CC_ protein genes, and AAE01546-AAE01630 represent the proteins they encode.
CC_ AAE01631-AAE01660 represent human secreted protein fragments or variants.
CC_ The secreted proteins and their genes are useful for preventing,
CC_ treating or ameliorating medical conditions, e.g., by protein or gene
CC_ therapy. Pathological conditions can be diagnosed by determining the
CC_ amount of the new protein in a sample or by determining the presence of
CC_ mutations in the new genes. Specific uses are described for each of the
CC_ 29 genes, based on the tissues in which they are most highly expressed,
CC_ and include developing products for the diagnosis or treatment of
CC_ proliferative disorders, cancer, tumours, foetal and developmental
CC_ abnormalities, haematopoietic disorders, diseases of the immune system,
CC_ AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC_ allergies, neurological disorders (e.g., Alzheimer's disease,
CC_ Parkinson's disease), cognitive disorders, schizophrenia, asthma,
```

```
CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC cardiovascular disorders, angiogenic disorders, kidney disorders,
CC gastrointestinal disorders, pregnancy-related disorders, endocrine
CC disorders, and infections. The proteins can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues, to identify their
CC cognate ligands or binding partners, and in chemotaxis, and can be used
CC as a food additive or preservative to modify storage properties.
CC Antibodies specific for a protein of the invention can be used in
CC alleviating symptoms associated with the disorders mentioned above, and
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC immunosorbent assay (ELISA). The present sequence represents a human
CC secreted protein of the invention.
XX
SQ_ Sequence 314 AA;

alignment_scores:
Quality: 147.50 Length: 163
Ratio: 2.049 Gaps: 10
Percent Similarity: 44.172 Percent Identity: 33.742

alignment_block:
US-09-462-480-2/rev x AAE01594 ..
Align seg 1/1 to: AAE01594 from: 1 to: 314

472 CGGCGCGGGTGGCGGGAAGTCTGTTCATTACGGGAGCTCACCAGTCG 423
|||||
126 ArgProGly***Pro..... 130
422 TCCTCTTCGTCGCGAGTCGT.....CCTC 400
|||||
131 .ProLeuProProSerSerSerGlyAsnAspGluThrProArgLeuProG 147

399 GTCGCTCTTTCACGCTCTCGCGAGCGGTGCGGCGCGACGACGCCG 350
|||||
147 InArgAsnLeu...SerLeuSerSerThrProProLeuProSerPro 162

349 GGCTGGTGGAGCGCGGATTGCGAACCTGGCCCATCGCTCCCGGACCC 300
|||||
163 Gly.....ArgSerGlyPro...LeuProProPr 171

299 ACGGAGCGGCGCACCGCTCACCGACGATCCGGCAACAGCGCGCGCAT 250
|||||
171 OProSerGluArg...ProProProValArgAspProProGlyArgS 187

249 CA.....CCGAGGGGGCAACCGCTTTTCGATCAGTCAGTACATCA 209
|||||
187 erGlyProLeuProProProValSerArgAsnGlySerThrSer 203

208 GCGCGTGGGGTCAACGACCCACCTCGGCA.....GGTAGCGAC 168
|||||
204 ArgAlaLeuProAlaThr...ProGlnLeuProSerArgSerGlyValAsp 219

167 TCGCGCGCGACAGCGCGCGCTGGGGCTGTATCCACCAGCCAG 118
|||||
220 SerProArgSerGlyProArgProProLeuProProAspArgPro..... 234

117 CGGATGTTTCACAGCGGAGCTGTGCGGAGCAGGCCCATCTGCGCGGCTT 68
|||||
234 ..... 234

67 CCTCGTGGGTGGTTCGCGCGCGCGGTGCGGCCACC 30
|||||
235 .....SerAlaGlyAlaProProProProSer 245

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:AAE54248
seq_documentation_block:
ID_ AAB54248 standard; Protein; 200 AA.
XX
```



(HYSE-) HYSEQ INC.

PA Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;  
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
PI Xue AJ, Yang Y, Wehrman T, Goodrich R;  
XX WPI; 2001-476283/51.  
DR N-PSDB; AAK52974.

XX Nucleic acids encoding polypeptides with cytokine-like activities,  
PT useful in diagnosis and gene therapy -  
PT Claim 20; Page 355-356; 6221pp; English.

XX The invention relates to polynucleotides (AAK51456-AAK53435) and the  
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
CC (AAM80020) are omitted as the relevant pages from the sequence listing  
CC were missing at the time of publication.

XX Sequence 1092 AA;

alignment\_scores:  
Quality: 141.50 Length: 210  
Ratio: 1.626 Gaps: 13  
Percent Similarity: 41.429 Percent Identity: 30.476

alignment\_block:

US-09-462-480-2/rev x AAM79841

Align seg 1/1 to: AAM79841 from: 1 to: 1092

469 CCGGGTGGCGCG..... 458  
||||:|||||  
52 ProAlatrpProArgHisProCysProProSerLeuThrProGlyProSe 68  
457 .....GGAAGTCTGTGTCATTACGGGAGCTACCCAGTCGTCCTC 418  
68 rMetAlaGlnGlyAlaMetArgPheCysSerGluGlyAspCysAlaIles 85  
417 TTCGTCCTCCAGTCGCTCGCTCTTCTTCAGCTCTCGCGAGCGGTG 368  
85 erProProArgCysProArgArgTrpLeu.....ProGluGlyProVal 99  
367 CCGGCGCGACACACCGCGGCTGTGGAGC.....CG 336  
100 ProGlnSerProProAlaSerMetTyrGlySerThrGlySerLeuLeuAr 116  
335 CCGGATTGCGAACCTGGCCCA.....TCGC 310  
116 gArGValAlaGlyProGlyProArgGlyArgGluLeuGlyArgValThrA 133  
309 TCCGGACCCACCGAGCGCGCCACCGTCACCGAGCATCCGCAACAG 260  
133 laProCysThrProLeuArgGlyProProSerProArgValAlaProSer 149  
259 CCG.....CCGGATCACCAGGGGCAACCG..... 233  
150 ProTrpAlaProSerSerProThrGlyGlnProProGlyAlaGlnSe 166  
232 .....GCTTTCGATCAGCTGAGACATCAGCGCGCTGCGG...GT 196  
166 rSerValValIlePheArgPheValGluLysAlaSer..ValArgProLe 182

195 CAACGACCCACCTGGCGGAGGTAGCGACTCC..... 165  
:|||||  
182 uAnGlyLeuProAlaProGlyGlyLeuSerArgSerTrpAspLeuGlyG 199  
164 .....GCGCGCAGCAGCGCGCGCGCGCTGGGGCCT..... 132  
:|||||:|||||  
199 lyValSerProProArgProThrProAlaLeuGlyProGlySerAsnArg 215  
131 .....GATCCA...CCAGCCAGCGGATG 112  
216 LysLeuArgLeuGluAlaSerThrSerAspProLeuProAlaArgGly.. 231  
111 GTTCGACGAGCGGACTGGTCGAGCAGCGCCATCTGCGCGCTTCCTCGT 62  
232 .....GlySerAlaLeuProGlySerArgAsnLeuValHisGly..... 244  
61 CGGCTGGGTGGCGCGCGCGGTGCGCGCC 33  
245 .....ProProAlaProPro 249

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1999.DAT.AAY24091

seq\_documentation\_block:

ID AAY24091 standard; Protein; 504 AA.

XX AC AAY24091;  
XX DT 08-SEP-1999 (first entry)  
XX DE Human Wiskott-Aldrich syndrome protein interacting protein.  
XX KW Human; Wiskott-Aldrich syndrome; Wiskott-Aldrich syndrome protein;  
XX KW WASP; WAS; WASP-interacting protein; WIP; actin.  
XX OS Homo sapiens.

XX WO9932628-A2.

XX PD 01-JUL-1999.

XX PF 22-DEC-1998; 98WO-US27501.

XX PR 23-SEP-1998; 98US-0101457.

XX PR 23-DEC-1997; 97US-0068533.

XX PA (CHIL-) CHILDRENS MEDICAL CENT.

XX PI Anton IM, Geha RS, Hartwig JH, Ramesh N;

XX DR WPI; 1999-418927/35.

XX DR N-PSDB; AAX83972.

XX PT A nucleic acid encoding a Wiskott-Aldrich syndrome protein  
XX PT (WASP)-interacting protein  
XX PS Claim 9; Fig 1; 50pp; English.

XX CC The present sequence represents a human Wiskott-Aldrich syndrome protein  
CC (WASP)-interacting protein (WIP). WIP or DNA encoding WIP can be  
CC administered to an individual, in sufficient quantity to alter actin  
CC content and/or the extent to which polymerisation occurs, and thus, to  
CC prevent or reduce (totally or partially) cytoskeletal abnormalities and  
CC other adverse effects. In particular, Wiskott-Aldrich syndrome can be  
CC treated or prevented in this manner.

XX SQ Sequence 504 AA;

alignment\_scores:  
Quality: 140.00 Length: 163  
Ratio: 2.000 Gaps: 9  
Percent Similarity: 42.945 Percent Identity: 32.515

```
alignment_block:
US-09-462-480-2/rev x AAY24091
Align seg 1/1 to: AAY24091 from: 1 to: 504

472 CGGCGCGGGTGGCGGGGAAGTCTGTTCATTCAGGGAGCTCACCAGTCG 423
|||||
315 ArgProGlyProPro..... 319
422 TCCTCTTCGTCCTCCAGTCGT.....CCTC 400
|||||
320 .ProLeuProSerSerSerGlyAsnAspGluThrProArgLeuProG 336
399 GTCGTCCTTCACGCTCCTCGCGAGCGGTGCGGGCGCGCAGACCCG 350
:|:|:| :|:|:| :|:|:| :|:|:| :|:|:|
336 lnArgAsnLeu...SerLeuSerSerThrProLeuProSerPro 351
349 GGCTGGTGGAGCCCGCGGATTGCCAACCTGGCCCATCGTCCCGGACCC 300
||| :|:| :|:| :|:| :|:|
352 Gly.....ArgSerGlyPro...LeuProProPr 360
299 ACCGAGCGGGCCACCGTCACCGACGATCCGGCAACAGCGCGCGCAT 250
|:|:|:|:|:|:| :|:|:|:|:|:| :|:|:|:|
360 ovalProSerGluArgProProProValArgAspProGlyArgS 377
249 CA.....CCGAGGGGCAACCGGCTTTCGATCAGCTGACACATCA 209
||| :|:|:| :|:|:| :|:|:|:|:|:|
377 erGlyProLeuProProProProValSerArgAsnGlySerThrSer 393
208 CGGCGTGGCGGTCAACGACCCACCTCGGCCA.....GTAGCGAC 168
||| :|:|:| :|:| :|:|
394 ArgAlaLeuProAlaThr..ProGlnLeuProSerArgSerGlyValAsp 409
167 TCCGCGCGCAGCGCGCGCGCGCTGGGCGCTGATCCACCGCCAG 118
||| :|:|:|:|:| :|:| :|:| :|:|
410 SerProArgSerGlyProArgProProLeuProProAspArgPro..... 424
117 CGGATGTTGACAGCGGACTGTGTCGCGAGCAGGCCCATCTGCGCGGCTT 68
424 ..... 424
67 CTCGTCGGTGGTTGCGCGCGCGGTGGCGCCACC 30
||||| :|:|:|:|:| :|:|:|:|
425 .....SerAlaGlyAlaProProProProProSer 435
```



OM of: US-09-462-480-2 to: Issued\_Patents\_AA:\* out\_format : pfs  
Date: Jul 22, 2002 1:24 AM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:  
-MODEL=frame+n2p.model -DEV=xlh  
-O=/cgn2\_1/USPTO.spool/US09462480/runat\_18072002.164418.19425/app\_query.fasta\_1.2850  
-DB-Issued\_Patents\_AA -QPWT=fastan -SUFFIX=ra1 -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000  
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-GAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsum62  
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09462480 -CGN1\_1\_18 -NCPD=6 -ICPD=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

Search information block:  
Query: US-09-462-480-2  
Query length: 524  
Database: Issued\_Patents\_AA:\*  
Database sequences: 231628  
Database length: 24425594  
Search time (sec): 53.850000

score\_list:

Sequence	Strd Orig	Zscore	EScore	Len	Documentation
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-08-818-112-114 +	713.00	1096.12	3.9e-54	368	
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-08-818-111-109 +	713.00	1096.12	3.9e-54	368	
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-09-056-556-114 +	713.00	1096.12	3.9e-54	368	
/cgn2_6/ptodata/2/1aa/5B_COMB.ppt:US-09-010-928B-2 +	128.00	181.30	0.0017	745	
/cgn2_6/ptodata/2/1aa/5B_COMB.ppt:US-09-010-928B-2 +	128.00	179.83	0.0018	870	
/cgn2_6/ptodata/2/1aa/5B_COMB.ppt:US-07-945-283-2 +	125.50	190.90	0.0021	191	
/cgn2_6/ptodata/2/1aa/5A_COMB.ppt:US-07-945-283-2 +	125.50	168.22	0.0036	1958	
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-09-413-814-78 +	125.00	175.04	0.0033	882	
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-09-232-468A-8 +	121.00	176.27	0.0061	404	
/cgn2_6/ptodata/2/1aa/5B_COMB.ppt:US-09-080-897-2 +	120.50	164.75	0.0087	1248	
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-09-323-735-2 +	120.50	164.75	0.0087	1248	
/cgn2_6/ptodata/2/1aa/6A_COMB.ppt:US-08-899-595-3 +	120.00	163.47	0.0097	1315	
/cgn2_6/ptodata/2/1aa/6A_COMB.ppt:US-08-483-533-41 +	118.50	173.62	0.0098	355	
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-09-283-471A-41 +	118.50	173.62	0.0098	355	
/cgn2_6/ptodata/2/1aa/PCTUS_COMB.ppt:PCT-US91-06532-3 +	118.50	173.62	0.0098	355	
/cgn2_6/ptodata/2/1aa/5B_COMB.ppt:US-09-010-928B-4 +	118.00	163.91	0.0134	907	
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-08-483-533-40 +	116.00	172.56	0.0151	264	
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-09-283-471A-40 +	116.00	172.56	0.0151	264	
/cgn2_6/ptodata/2/1aa/6A_COMB.ppt:US-08-586-165-4 +	115.00	168.08	0.0198	359	
/cgn2_6/ptodata/2/1aa/5A_COMB.ppt:US-07-928-611-13 +	112.00	166.24	0.0336	267	
/cgn2_6/ptodata/2/1aa/5B_COMB.ppt:US-08-487-811A-13 +	112.00	166.24	0.0336	267	
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-09-060-694-13 +	112.00	166.24	0.0336	267	
/cgn2_6/ptodata/2/1aa/PCTUS_COMB.ppt:PCT-US93-07370-13 +	112.00	166.24	0.0336	267	
/cgn2_6/ptodata/2/1aa/5A_COMB.ppt:US-08-056-051-6 +	112.00	160.91	0.0381	467	
/cgn2_6/ptodata/2/1aa/5A_COMB.ppt:US-07-928-611-22 +	112.00	160.91	0.0381	467	
/cgn2_6/ptodata/2/1aa/5B_COMB.ppt:US-08-487-811A-22 +	112.00	160.91	0.0381	467	
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-09-060-694-22 +	112.00	160.91	0.0381	467	
/cgn2_6/ptodata/2/1aa/PCTUS_COMB.ppt:PCT-US93-07370-22 +	112.00	160.91	0.0381	467	
/cgn2_6/ptodata/2/1aa/PCTUS_COMB.ppt:PCT-US91-06532-2 +	111.50	165.61	0.0370	263	
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-09-041-886-23 +	111.00	150.49	0.0571	1185	
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-08-764-870-7 +	110.50	157.47	0.0527	525	
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-08-980-115-7 +	110.50	157.47	0.0527	525	
/cgn2_6/ptodata/2/1aa/backfiles1.ppt:5168049-3 +	110.00	166.73	0.0461	183	
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-08-483-533-26 +	110.00	163.46	0.0497	258	
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-09-283-471A-26 +	110.00	163.46	0.0497	258	
/cgn2_6/ptodata/2/1aa/5A_COMB.ppt:US-08-131-365B-38 +	110.00	149.25	0.0692	1147	
/cgn2_6/ptodata/2/1aa/5B_COMB.ppt:US-08-668-123-38 +	110.00	149.25	0.0692	1147	
/cgn2_6/ptodata/2/1aa/5B_COMB.ppt:US-08-728-323A-2 +	110.00	149.13	0.0694	1162	
/cgn2_6/ptodata/2/1aa/6B_COMB.ppt:US-09-423-890-8 +	110.00	146.74	0.0733	1493	
/cgn2_6/ptodata/2/1aa/backfiles1.ppt:5171843-9 +	109.50	159.05	0.0598	378	

/cgn2\_6/ptodata/2/1aa/6B\_COMB.ppt:US-09-199-637A-351 - 109.00 156.67 0.0685 447  
/cgn2\_6/ptodata/2/1aa/6B\_COMB.ppt:US-08-556-978B-59 - 109.00 155.74 0.0700 493  
/cgn2\_6/ptodata/2/1aa/6B\_COMB.ppt:US-09-199-637A-277 + 109.00 154.56 0.0720 551  
/cgn2\_6/ptodata/2/1aa/5A\_COMB.ppt:US-07-945-283-2 - 109.00 142.61 0.0950 1958  
/cgn2\_6/ptodata/2/1aa/backfiles1.ppt:5171843-7 + 108.50 162.14 0.0655 232  
seq\_name: /cgn2\_6/ptodata/2/1aa/6B\_COMB.ppt:US-08-818-112-114  
seq\_documentation\_block:  
; Sequence 114. Application US/08818112  
; Patent No. 6290969  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Campos-Neto, Antonio  
; APPLICANT: Houghton, Raymond  
; APPLICANT: Vedvick, Thomas S.  
; APPLICANT: Twardzik, Daniel R.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY  
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS  
; NUMBER OF SEQUENCES: 153  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA: US/08/818,112  
; APPLICATION NUMBER: 13-MAR-1997  
; FILING DATE: 13-MAR-1997  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.411C6  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 114:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 368 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-818-112-114  
alignment\_scores:  
Quality: 713.00 Length: 143  
Ratio: 5.057 Gaps: 0  
Percent Similarity: 98.601 Percent Identity: 96.503  
alignment\_block:  
US-09-462-480-2 x US-08-818-112-114 ..  
Align seg 1/1 to: US-08-818-112-114 from: 1 to: 368  
1 CTGCAGAGGTGACGTGTTTTCAGCCAGTGGCGGCACCGCGCGG 50  
|||||  
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGly 242  
51 CAACCCAGCCGACGAGGAGCGCGAGATGGCGCTCTCGGCACGATC 100  
242 YasnProAlaAspGluAlaAlaGlnMetGlyLeuLeuGlyThrSerP 259  
101 CGCTGTGCAACCATCCGCTGCTGCTGATCAGGCCCGCGCGCGCG 150



; TITLE OF INVENTION: EXTREMELY ELASTIC SPIDER SILK PROTEIN AND DNA  
; TITLE OF INVENTION: CODING THEREFOR

10

```
36 aAlaAlaAlaGlyGlyProLeuArgTyrGlyArgAlaValArgAlaArgG 53
52 AACCC.....AGCCGACGAGAACCCGCCGAGATGGCGCTGCTGGCAC 95
   |||  ::||| |||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
53 lyProArgAspAlaArgArgGlyAla.....AlaProGlyArgGly 66
   ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
96 CAGTCCGCTGTCGAACCATCGCTGGTGGATCAGGCCCGCCAGCGCG 145
   ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
67 AlaArgSerAlaThrProGlyAlaGly.....AlaProAlaArgG1 80
   ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
146 G.....CGCGGCGCTGCTGGCGCG 165
   ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
80 yHisArgAlaArgAlaProAlaProArgArgArgGlyProAlaAlaArgG 97
   ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
166 GAGTCGCTACCTGGCGCAGGTGGTGTGACCCGCGCAGCGCTGATGTC 215
   ||
97 ly..... 97

216 TCAGCTGATCGAAAAAGCGGTTGCCCTCGGTGATGCCGCGGTGTG 265
   |||  |||  |||  ::|||  ::|||  ::|||  ::|||  ::|||
98 .....GlyArgGlyGlyGlyProArgAlaGlyAlaLe 108

266 CGGATCTCGTGCAGCGGTGGCGCC.....GCTCCGGTGGTCCG 306
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
108 uArgAlaGlyGly.ArgGlyGlyAlaArgGlyProAlaGluGlyAla 124

307 GGAGCGATGGCGAGGTTGCCAATCCGCGCGCTCCACCGCCGCGG... 353
   |||||  |||  |||||  |||||  |||||  |||||  |||||  |||||
125 GlyAlaAlaGly.GlyValArgLeuProAlaAlaProPro.ProGlyArg 140

354 TCTGTCTCGCGCGCGCTCGCGCAGGAGCGTGAAGAAGCAGCAGG 403
   ::|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
141 GlyGlyArgAlaAlaArgProAspProGlyLeuArgArgAlaGlyAl 157

404 AGGACTGGGACGAGCAGCAGTGGTGAAGTCCCGTAATGAC 446
   ::|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
157 aAspAlaGlyArgAspAlaArgArgProGluValArgArgAsp 171
```

seq\_name: /cgn2\_6/ptodata/2/iaa/5A\_COMB pep:US-07-945-283-2

```
seq_documentation block:
; Sequence 2, Application US/07945283
; Patent No. 5352596
; GENERAL INFORMATION:
; APPLICANT: Cheung, Andrew K.
; APPLICANT: Wesley, Ronald D.
; TITLE OF INVENTION: Pseudorabies Virus Deletion Mutants
; TITLE OF INVENTION: Involving The EP0 and LIT Genes
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis P. Ribando
; STREET: 1815 No. 5352596th University Street
; CITY: Peoria
; STATE: IL
; COUNTRY: USA
; ZIP: 61604
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/945,283
; FILING DATE: 19920911
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Ribando, Curtis P
; REGISTRATION NUMBER: 27976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 309-685-4011 ext.513
; TELEFAX: 309-685-4128
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
```

```
; LENGTH: 1958 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-945-283-2

alignment_scores:
  Quality: 125.50      Length: 238
  Ratio: 1.195        Gaps: 16
  Percent Similarity: 44.118      Percent Identity: 31.092

alignment_block:
US-09-462-480-2 x US-07-945-283-2 ..
Align seg 1/1 to: US-07-945-283-2 from: 1 to: 1958

31 GTGGCGGCACCGGC...GGCGGCAACCCA..... 57
   |||||  |||  |||||  ::|||
1053 ValGlyGlyGluGlyArgGlyGlyGlyProArgArgValGlyLeuAlaG1 1069
   58 .....GCCGACGAGGAGCGCGCAGATGGGCTGCTGGCACCAGTC 100
   |||  |||  |||  ::|||  ::|||  ::|||  ::|||  ::|||
1069 yArgAspAlaAlaGluAlaValAlaGlyArgGlyValLeuGlyHisGlyP 1086
   101 CG...CTGTGCAACCATCCG.....CTGGCTGGTGA..... 129
   ||  ::|||  |||  |||||  |||||  |||||  |||||  |||||
1086 roGluArgAlaProGluProValValLeuGlyGlyGlyGlyGlyGly 1102
   130 .....TCAGGCCCC...AGCGCGG 146
   |||||  |||  |||||  |||  |||||  |||  |||  |||
1103 GlyGlyHisGluArgGlySerGlyValArgSerGlyProGluSerGluG1 1119
   147 CGCGGCGCTG..... 156
   |||||  |||
1119 yAlaAlaLeuAlaProGlyProProValLeuPheValValAlaValAlaV 1136
   157 .....CTGCGCGCGGAG.....TCG 171
   ::|||  |||||  ::|||
1136 alAlaValProAlaGluGlyArgAlaGlyGluProLeuValLeuAla 1152
   172 CTACTGGCGCAGGTGGTGTGACCGCACGCGCGCTGTGTCTCAGCT 221
   ::|||  |||||  ::|||  |||  |||||  ::|||  ::|||
1153 ValProGlyAlaAlaGly.....ProGlyArgAlaAlaLe 1164
   222 GATCGAAAAGCGGTTGCCCTCGGTGATGCCGCGCTGTGTCCCGGAT 271
   |||  ::|||  ::|||  |||  |||||  ::|||  ::|||
1164 uLeuLeuAlaProLeuGlyArgTrpValArg...AlaGlyGlyGlyGlyA 1180
   272 CGTGGTGACGCGGTGGCGCGCTCGGTGGTGGTCCGCGGAGCG..... 312
   ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
1180 laGlyValAlaGlyGlyAlaGlyGluAlaGlyLeuGlyAlaGlyAlaGly 1196
   313 .....ATGGCCAGGGTTCCAAATCCGCGCGCTCCAC.CAGCCCGGTC 355
   |||  |||  ::|||  ::|||  ::|||  ::|||  ::|||
1197 LeuGlyAlaGlyAlaGlyLeuGlyAlaGlyGlyAlaGlyGlyProGlyAl 1213
   356 TGGTCGCGCGCGCACCCTCGCGCAGGAGCGGTGAGAGAGACA..... 398
   ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
1213 aGlyGluAlaGly.....GlyGlyAlaArgArgArgArgArgArgA 1227
   399 .....CGAGGACGACTGGGA 413
   |||||  ::|||
1227 rgTrpAspAspGluAlaGlyLeuLeuGlyProGluArgGlyGlnAlaGly 1243
   414 CGA.....AGAGCAGCAGTGGTGAAGTCCCGTAA 442
   |||  |||||  |||  ::|||
1244 ArgGlyLeuArgGlyProGlyProArgGlyGlyLeuGlyGlyProGly.. 1259
   443 TGACACAGACTTCCCGCGCACCAGGCGGCGGAGACCTTGCACACATTTTG 492
   |||||  |||||  |||||
1260 .....ProGlyHisValGlyArg.....G 1266
```

493 GCGAGGAGGTAAA 506  
|||||  
1266 lyglucluylarg 1270

seq\_name: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.US-09-413-814-78

seq\_documentation\_block:  
; Sequence 78, Application US/09413814  
; Patent No. 6225064  
; GENERAL INFORMATION:  
; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH  
; APPLICANT: Bristol-Myers Squibb, Co.  
; APPLICANT: Beyer, Stefan  
; APPLICANT: Bloecker, Helmut  
; APPLICANT: Brandt, Petra  
; APPLICANT: Cino, Paul M  
; APPLICANT: Dougherty, Brian A  
; APPLICANT: Goldberg, Steven L  
; APPLICANT: Hofle, Gerhard  
; APPLICANT: Mueller, Joachim  
; APPLICANT: Reichenbach, Hans  
; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or  
; TITLE OF INVENTION: heteropolymers of polyketide compounds  
; FILE REFERENCE: PCT/US 99/23535  
; CURRENT APPLICATION NUMBER: US/09/413,814  
; CURRENT FILING DATE: 1999-10-07  
; EARLIER APPLICATION NUMBER: DE 198 46 493.2  
; EARLIER FILING DATE: 1998-10-09  
; NUMBER OF SEQ ID NOS: 107  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 78  
; LENGTH: 882  
; TYPE: PRT  
; ORGANISM: Sorangium cellulosum  
US-09-413-814-78

alignment\_scores:  
Quality: 125.00 Length: 138  
Ratio: 1.786 Gaps: 10  
Percent Similarity: 50.725 Percent Identity: 37.681

alignment\_block:  
US-09-462-480-2/rev x US-09-413-814-78 ..

Align seg 1/1 to: US-09-413-814-78 from: 1 to: 882

424 GTCTCTTCCTCCAGTCGTCGTCGTCCTTCTTCACGCTCTCGCG... 377  
||||| |||||||  
347 ArgProThrArgProSerArgProArg... 360  
376 ....CGAGCGGTGCGGCGGCGACCCAGCCGCGGCTGGTGAGCGCCGGA 331  
||| |||||||  
360 gArgArgSerProAlaArg..... 367  
330 TTGCGAACCTGCGCCATCGCTCCCGACCCAGCGGCGCGCACCG 281  
:::|||||  
368 .....ThrProGlyProArgSerPro.....ArgArgPro 378  
280 TCACCGAGATCCGG.....CAACAGCGCGCGCATCACC 246  
|||||  
379 SerProSerAlaArgSerProAlaIleTrpGluArgProProArgArgPr 395  
245 GAGGGGCAACCGCTTTTCGATCAGCTGAGACATCAGCGCGTGGCGGT 196  
|||||  
395 oArgAspArgProArgArg..LeuLeuArgProAlaArgArgLeuAla 411  
195 CAACAGCCACCT.....GCGCCAGGTAGCAGCTCCCGCGCGCAGCAGGC 152  
:::|||||  
412 ArgArg..AspProGlyAspValProAspProAlaAlaArgArg.. 427  
151 CCGCGCCGCGCTGGCGCTGATCCACCGACCGCGGATGTTTCACAGC 102  
||||| ||| ||| |||

428 ..AlaProAlaAlaGlyAlaLeuProGlyAlaAspGlyArgGlyAlaArg 443  
101 GGACTGGTGGCGAGCAGCGCCATCTCGCGGCTTCTCGCTCGCTGGTT 52  
||| |||  
444 Gly.....ProGlyArgArgAlaProArgArgGlyAlaAlaAlaAr 458

51 GCGCGCGCGC 42  
|||||  
458 gValProPro 461

seq\_name: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.US-09-232-468A-8

seq\_documentation\_block:  
; Sequence 8, Application US/09232468A  
; Patent No. 6207165  
; GENERAL INFORMATION:  
; APPLICANT: AUDONNET et al.  
; TITLE OF INVENTION: POLYNUCLEOTIDE VACCINE FORMULA AGAINST PORCINE  
; TITLE OF INVENTION: REPRODUCTIVE AND RESPIRATORY PATHOLOGIES  
; FILE REFERENCE: 454313-2230  
; CURRENT APPLICATION NUMBER: US/09/232,468A  
; CURRENT FILING DATE: 1999-01-05  
; NUMBER OF SEQ ID NOS: 54  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 8  
; LENGTH: 404  
; TYPE: PRT  
; ORGANISM: Aujeszky's Disease Virus (NIA3 Strain)  
US-09-232-468A-8

alignment\_scores:  
Quality: 121.00 Length: 166  
Ratio: 1.424 Gaps: 7  
Percent Similarity: 51.205 Percent Identity: 30.723

alignment\_block:  
US-09-462-480-2/rev x US-09-232-468A-8 ..

Align seg 1/1 to: US-09-232-468A-8 from: 1 to: 404

472 GCGCGCGGTGGCGGGAAGTCTGTGTTCATTACGGAG...CTCACCAG 426  
||||| |||  
193 ArgProArgGlyProAlaProHisValGlnValArgArgValLeuGluAr 209  
425 TCCTCTCTTCCTCCAGTCGTCGTCGTCCTTCTTCACGCTCTCGCGC 376  
||| |||  
209 gArgGlnLeuGlnAlaGlyArgGlyArgAspAlaIle.....ProAspA 224  
375 GAGCGGTGCGCGCGACGACCGCGGCTGGTGAGCGCGCGGATTGCG 326  
::: ||| |||  
224 laValLeuProAlaAlaProAlaProGlyGlyGlyLeuLeuValPro 240  
325 AACCTTGGCCATCGCTCCCGGACCCAGCGGCGCGCCACCGCTCACC 276  
::: ||| |||  
241 GlnGluArgProAspAlaProAlaGlyProArgArgHis..... 254  
275 GACGATCGGCAACAGCGCGCGCA.....TCAC 247  
::: ||| |||  
255 .AlaValArgHisArgProArgAlaAlaLeuGlyGlyLeuAlaGluAlaP 271  
246 CGAGGGGCAACCGCTTTTCGATCAGCTGAGACATCAGCGCGCTGCGGG 197  
||| |||  
271 roAlaProAlaProAlaProAlaGluAlaArgAlaArgProGly 287  
196 TCACAGCCACCTGCGCGCAGGTAGCAGCTCCCGCGCGCAGCGCCCGC 147  
::: ||| |||  
288 AspAlaArgAlaProArgProProAlaAlaAlaGlyAspAlaGlyProAr 304  
146 CCCCGCTGGGCGCTGATCCACCGACCGCGGATGTTTCGACAGCGGACT 97  
||| |||  
304 gArgArg.....GlyProHisAlaAlaThrProGluA 316

```

96 GGTGCGGAGCAGGC.....CCATCTGGCGCGCTTCTCTCGTCGG 59
||||| |||:|||||
316 laArgAspAlaAlaProLeuArgProAlaGlyArgAlaGlnArg 332
||||| |||:|||||
58 CTGGTTCCCGCGCGGTGGCGCCACCTGGCTGACACAGAGGT 13
|||:||||| |||:|||||
333 Val.AlaAlaAlaArgGlyAlaValProAlaAlaAspProArg 347
|||:||||| |||:|||||

```

```

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-09-080-897-2

```

```

seq_documentation_block:
; Sequence 2, Application US/09080897
; Patent No. 5985574
; GENERAL INFORMATION:
; APPLICANT: King, Mary-Claire
; APPLICANT: Lynch, Eric D.
; APPLICANT: Lee, Ming
; APPLICANT: Morrow, Jan E.
; APPLICANT: Welcsh, Piri L.
; APPLICANT: Leon, Pedro E.
; TITLE OF INVENTION: Modulators of Actin
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 75 DENISE DRIVE
; CITY: HILLSBOROUGH
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94010
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/080,897
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UW97-001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 343-4341
; TELEFAX: (650) 343-4342
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1248 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-080-897-2

```

```

alignment_scores:
  Quality: 120.50      Length: 149
  Ratio: 1.826        Gaps: 8
Percent Similarity: 44.295 Percent Identity: 33.557

```

```

alignment_block:

```

```

US-09-462-480-2/rev x US-09-080-897-2

```

```

Align seg 1/1 to: US-09-080-897-2 from: 1 to: 1248

```

```

448 TTGTCATTACGGAGCTCACAGTCGTCTCTTCTGCCAGTCGCTCG 399
||||| |||:|||||
556 LeuSerAlaAlaAlaIleThrValProProSerValProSerArg..... 570
||||| |||:|||||
398 TCGTCTTCTTACGCTCTCGCGAGCGGTGCGCGCGAGCAGACCGG 349
||||| |||:|||||
571 .....AlaProValProProAlaProLeuProG 581
348 GCTGTGAGCGCGCGGATTGCGAACCTGCGCCCATCGTCTCCCGACCCA 299

```

```

||| ||| ||| ||| |||
581 ly.....AspSerGlyThrIleIleProPro 590
||||| |||:|||||
298 CCG.....GAGCGCGCCACCCGTCACCGAGGATCCGGCAACAGCC 258
||||| |||:|||||
591 ProAlaProGlyAspSerThrProProProProProProProPr 607
||||| |||:|||||
257 GCCGCGATCACCGAG.....GGGCAACCGGCTTTTCGATCAGCTGAGACA 212
||||| |||:|||||
607 oProProProLeuProGlyGlyThrAlaIleSerProProProL 624
||||| |||:|||||
211 TCAGCGCGTCTCGGTCAACGCCACCTGCGCCA..... 177
||||| |||:|||||
624 euSerGlyAspAlaThrIleProProProProLeuProLeuGlyVal 640
||||| |||:|||||
176 GGTAGCGACTCGCGCGCAGCAGGCC...CGCCCGCGCTGGGCGCTGA 130
||||| |||:|||||
641 GlyIleProProSerSerLeuProGlyGlyThrAlaIleProProPr 657
||||| |||:|||||
129 TCCACCGCCAGCGGATGTTTCGACAGCGGACTGTGCGCGAGCGCCA 80
||||| |||:|||||
657 oProProLeuProGly.....SerAlaArgIleProProProPro 670
||||| |||:|||||
79 TCTCGCGCGTCTCTCGTCTCGGTGCGCGCGCGCGTGGCGCGCC 33
||||| |||:|||||
671 ..ProProLeuProGlySerAlaGlyIleProProProProProPro 685
||||| |||:|||||
seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-323-735-2

```

```

seq_documentation_block:
; Sequence 2, Application US/09323735
; Patent No. 6197932
; GENERAL INFORMATION:
; APPLICANT: King, Mary-Claire
; APPLICANT: Lynch, Eric D.
; APPLICANT: Lee, Ming
; APPLICANT: Morrow, Jan E.
; APPLICANT: Welcsh, Piri L.
; APPLICANT: Leon, Pedro E.
; TITLE OF INVENTION: Modulators of Actin
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 75 DENISE DRIVE
; CITY: HILLSBOROUGH
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94010
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/323,735
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/080,897
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UW97-001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 343-4341
; TELEFAX: (650) 343-4342
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1248 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

```

US-09-323-735-2

```
alignment_scores:
  Quality: 120.50      Length: 149
  Ratio: 1.826         Gaps: 8
  Percent Similarity: 44.295      Percent Identity: 33.557

alignment_block:
US-09-462-480-2/rev x US-09-323-735-2      ..

Align seg 1/1 to: US-09-323-735-2 from: 1 to: 1248

448 TTGTCATTACGGAGCTCACCAGTCGTCTCTTCGTCCAGTCGTCGTCG 399
|||||
556 LeuSerAlaAlaAlaIleThrValProProSerValProSerArg..... 570

398 TCGTCTTCTTTCACGCTCTCGCGGAGCGGTGCGCGGCGGACGACCGG 349
|||||
571 .....AlaProValProProAlaProProLeuProG 581

348 GCTGTGTGAGCGCGGATTCGGAACCTGCGCCATCGCTCCGCGACCCA 299
||
581 ly.....AspSerGlyThrIleIleProProPro 590

298 CCG.....GAGCGCGCCACCGTCACCGAGCATCCGCGCAACGCC 258
|||
591 ProAlaProGlyAspSerThrThrProProProProProProProPr 607

257 GCGGCATCACCAG.....GGGCAACCGGTTTTCGATCAGTCAGTACACA 212
||||
607 oProProProProLeuProGlyGlyThrAlaIleSerProProProL 624

211 TCAGCGCGGTGCGGTCAACGACCCACCTCGCGCA..... 177
:|||||
624 euSerGlyAspAlaThrIleProProProProProProLeuProGlyVal 640

176 GGTAGCAGCTCCGCGCGCAGCAGGCC...GCGCGCGGCTGGGGCCTGA 130
|||
641 GlyIleProSerProSerSerLeuProGlyGlyThrAlaIleProProPr 657

129 TCCACCGACGAGGATGTTTCGACAGCGGACTGTCGCGGAGCGGCCA 80
|||||
657 oProProLeuProGly.....SerAlaArgIleProProProPro. 670

79 TCTGCGCGCTTCTCTCGCGGTGGTTCGCGCGCGGTGCGCGGCC 33
:|||||
671 ..ProProLeuProGlySerAlaGlyIleProProProProProPro 685

seq_name: /cgn2_5/ptodata/2/iaa/6A_COMB.pep:US-08-899-595-3
```

```
seq_documentation_block:
; Sequence 3, Application US/08899595
; Patent No. 6111072
; GENERAL INFORMATION:
; APPLICANT: Narumiya, Shuh
; APPLICANT: Takahashi, No. 6111072uaki
; TITLE OF INVENTION: RHO TARGET PROTEIN HUMAN MDIA AND GENE
; TITLE OF INVENTION: ENCODING SAME
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
```

```
; APPLICATION NUMBER: US/08/899,595
; FILING DATE: 24-JUL-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-242701
; FILING DATE: 26-AUG-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 9-90170
; FILING DATE: 25-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Stephen A. Bent
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 049441/0112
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1315 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-899-595-3

alignment_scores:
  Quality: 120.00      Length: 156
  Ratio: 1.714         Gaps: 8
  Percent Similarity: 44.872      Percent Identity: 32.051

alignment_block:
US-09-462-480-2/rev x US-08-899-595-3      ..

Align seg 1/1 to: US-08-899-595-3 from: 1 to: 1315

448 TTGTCATTACGGAGCTCACCAGTCGTCTCTTCGTCCAGTCGTCGTCG 399
|||||
608 LeuSerAlaAlaAlaIleThrValProProSerValProSerArg..... 622

398 TCGTCTTCTTTCACGCTCT..... 380
|||||
623 .....AlaProValProProAlaProProLeuProGlyAsps 635

379 ..GCGCGAGCGTGC CGCGCGAGCAGACCGCGGTGTCGCGCGCGG 332
:|:
635 erGlyThrIleIleProProProProAlaProGly.....AspSer 648

331 ATTGCGAACCTGGCCCATCGCTCCGCGACCCACCGGAGCGCGCGCACCC 282
:|:
649 ThrThrProProProProProProProProProProProProLeuPr 665

281 GTCACCGACGATCGGCAACAGCGCGGCATCACCGAGGGGCAACCGG 232
|:|:
665 oGlyGlyValCysIleSerSerProProSerLeuPro..GlyGlyThrAl 681

231 TTTTTCGATCAGTCAGACATCAGCGCGTGC GGTCACAGCCACCTG 182
:|:
681 aileSerProProProProLeuSerGlyAspAlaThrIleProProPr 698

181 CGCCA.....GGTAGCGACTCCGCGCGGACGAGGCC... 150
|||
698 roProLeuProGlyGlyValGlyIleProSerProSerSerLeuProGly 714

149 GCGCGCGCGTGGGCGCTGATCCACCGACCGAGCGATGGTTCGACAGCGG 100
:|:
715 GlyThrAlaIleProProProProProProProGly.....SerAl 728

99 ACTGGTCCGCGAGCGCCCATCTCGCGCGCTTCCTCGTCGTCGTCGTCG 50
:|:
728 aArgIleProProProPro...ProProLeuProGlySerAlaGlyIlePr 744

49 CGCGCGCGGTCCCGCCC 33
```



```
||||| 744 roProProProPro 749
seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-08-483-533-41
seq_documentation_block:
; Sequence 41, Application US/08483533
; Patent No. 6172047
; GENERAL INFORMATION:
; APPLICANT: Roizman, Bernard
; TITLE OF INVENTION: Method for Treating Tumorigenic
; DISEASES
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,533
; FILING DATE: 07-MAR-95
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/419,853
; FILING DATE: 11-APR-95
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/861,233
; FILING DATE: 31-MAR-92
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, James P.
; REGISTRATION NUMBER: 28,491
; REFERENCE/DOCKET NUMBER: 28097/32742
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELETYPE: 25-3856
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-483-533-41

alignment_scores:
Quality: 118.50 Length: 222
Ratio: 1.411 Gaps: 9
Percent Similarity: 37.838 Percent Identity: 26.126

alignment_block:
US-09-462-480-2/rev x US-08-483-533-41
Align seg 1/1 to: US-08-483-533-41 from: 1 to: 355
440 ACGGAGCT.....CACCAGTCGTCCTTCGTCCTCCAGTCGTCCTC 400
|||||
19 ThrGlyAlaValProThrAlaGlnSerGlnValThrSerThrProAsnSe 35
399 GTCGCTCTCTCAGCGCTCCTCGCGGAGCGGTGCGCGCGCGACGACCG 350
|||||
35 rGluProAlaValArgSerAlaProAlaAlaProProProProA 52
349 GGCTGGTGAGCGCGCGGATTGC..... 327
```

```
:::
52 laGlyGlyProProSerCysSerLeuLeuLeuArgGlnTrpLeuHis 68
|||||:::
326 .....GACCCTGGCCCATCGC 310
:::|||||:::
69 ValProGluSerAlaSerAspAspAspAspTrpProAspSe 85
|||||:::
309 TCCC.....GGACCCACCGGAGCGCG... 288
:::|||||:::|||||
85 rProProGluSerAlaProGluAlaArgProThrAlaAlaProA 102
287 ..CACCCGTCACGACGATCCGGCA.....ACA 261
|||||
102 rgProGlyProHisArgProAlaTrpAlaArgGlyAlaGlyLeuThr 118
260 GCCCGCGCATCACCGAGGGGCAACCGCTTTTCGATCAGCTGAGACAT 211
|||||:::|||||:::|||||
119 .ProProProProAlaProSerAlaPheArgAlaSerProS 135
210 CAGCGCGTGGGTCAACGACCCAGCTGGCGGAGGTAGGACTCCGCGC 161
|||||:::|||||
135 erAlaCysAlaSerProArgSerThrTrpArgAlaCysAlaCysAspAla 151
160 GCAGCAGCG..... 152
|||||
152 ArgAlaGlyArgGlyArgSerProProArgProProArgProProAr 168
151 .....CCGCGCGCGCTGGGCGCTGATCCACGACGCGCGG 115
|||||
168 gProProArgProProArgProProArgGlyCysAlaSerArgProThrS 185
114 ATGTTTCGACAGCGGACTGGTGGCGAGCGCGCCATCTGCG..... 74
:::|||||
185 erGlyCysAlaThrTrpSerGlyProArgProProAlaTrpArgAla 201
74 ..... 74
202 AlaAlaArgGlyProAlaSerGlyProThrGlyLeuGlySerGlyAlaGl 218
73 .....CGCTTCTCGTGGTGGTGGTGG.....CCGCGCGCGG 41
|||||
218 yTrpArgArgProArgSerSerGlyArgAlaTrpGlyProArgProV 235
40 TCGCGCGCACCTGG 27
|||||
235 alProGlyProTrp 239
seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-09-283-471A-41
seq_documentation_block:
; Sequence 41, Application US/09283471A
; Patent No. 6340673
; GENERAL INFORMATION:
; APPLICANT: Roizman, Bernard
; APPLICANT: Chou, Joany
; TITLE OF INVENTION: Method For Treating Tumorigenic Diseases
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/283,471A
; FILING DATE: 04-APR-1999
; CLASSIFICATION: 514
```

;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 07/861,233  
;; FILING DATE: 31-MAR-1992  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/419,853  
;; FILING DATE: 11-APR-1995  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/483,533  
;; FILING DATE: 07-JUN-1995  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Zeller, James P.  
;; REGISTRATION NUMBER: 28,491  
;; REFERENCE/DOCKET NUMBER: 27373/32742A  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 312/474-6300  
;; TELEFAX: 312/474-0448  
;; TELEX: 25-3856  
;; INFORMATION FOR SEQ ID NO: 41:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 355 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; US-09-283-471A-41

alignment\_scores:  
Quality: 118.50 Length: 222  
Ratio: 1.411 Gaps: 9  
Percent Similarity: 37.838 Percent Identity: 26.126

## alignment\_block:

US-09-462-480-2/rev x US-09-283-471A-41 ..

Align seg 1/1 to: US-09-283-471A-41 from: 1 to: 355

440 ACGGAGCT.....CACAGTCGTCCTGTCGCCAGTCGTCCTC 400  
|||||  
19 ThrGlyAlaValProThrAlaGlnSerGlnValThrSerThrProAsnse 35  
399 GTCGTCCTTCACGCTCCTCGCGGAGCGGTGCGCGCGCGACGACCG 350  
|||  
35 rGluProAlaValArgSerAlaProAlaAlaAlaProProProPro 52  
|||  
349 GCCTGGTGGAGCCCGCGATTGC..... 327  
|||  
52 laGlyGlyProProProSerCysSerLeuLeuLeuLeuArgGlnTrpLeuHis 68  
|||  
326 .....GAACCTGGGCCCATCGC 310  
|||  
69 ValProGluSerAlaSerAspAspAspAspAspTrpProAspse 85  
|||  
309 TCCC.....GGACCCACCGGAGCGCG..... 288  
|||  
85 rProProGluSerAlaProGluAlaArgProThrAlaAlaAlaProA 102  
|||  
287 ..CCACCCGTCACGACGATCCGSCA.....ACA 261  
|||||  
102 rgProGlyProHisArgProAlaTrpAlaArgGlyAlaGlyLeuThr 118  
|||  
260 GCCCGCGCATCACCAGGGGGGAACCGCTTTTCGATCAGCTGAGACAT 211  
|||||  
119 .ProProThrProProArgAlaProSerAlaPheArgArgAlaSerProS 135  
|||||  
210 CAGCGCGCTCGGTCAACGACCCACCTGCGCCAGGTAGGACTCCGCGC 161  
|||||  
135 erAlaCysAlaSerProArgSerThrTrpArgAlaCysAlaCysAspAla 151  
|||||  
160 GCAGCAGGC..... 152  
|||||  
152 ArgAlaGlyArgGlyArgArgSerProProArgProProArgProProAr 168

151 .....CCGCGCCCGCGCTGGGCGCTGATCCACGACGACGG 115  
|||||  
168 gProProArgProProArgProProArgGlyCysAlaSerArgProThrS 185  
|||  
114 ATGGTTTCGACAGCGGACTGGTGGCCGACGAGCCCATCTGCG..... 74  
:||||:||||: ||| ||| |||  
185 erGlyCysAlaThrTrpTrpSerGlyProArgProProAlaTrpArgAla 201  
74 ..... 74  
202 AlaAlaArgGlyProAlaSerGlyProThrGlyLeuGlySerGlyAlaG1 218  
73 ....CGGCTTCCTCGTCGCGTGGGTG.....CCGCGCGCGG 41  
||| ||||||| :||| ||| |||  
218 yTrpArgArgProArgSerSerGlyArgAlaTrpGlyProArgProV 235  
40 TGC GCGCCCACTGG 27  
||||| |||  
235 alProGlyProTrp 239

seq\_name: /cgn2\_6/ptodata/2/1aa/PCTUS\_COMB.pep:PCT-US91-06532-3

seq\_documentation\_block:  
;; Sequence 3, Application PC/TUS9106532  
;; GENERAL INFORMATION:  
;; APPLICANT: Roizman, Bernard  
;; TITLE OF INVENTION: Recombinant Herpes Simplex Viruses  
;; TITLE OF INVENTION: Vaccines and Methods  
;; NUMBER OF SEQUENCES: 8  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &  
;; ADDRESS: Bicknell  
;; STREET: Two First National Plaza Suite 2100  
;; CITY: Chicago  
;; STATE: Illinois  
;; COUNTRY: USA  
;; ZIP: 60603  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: PCT/US91/06532  
;; FILING DATE: 19910910  
;; CLASSIFICATION: 424  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Gruber, Lewis S.  
;; REGISTRATION NUMBER: 30,060  
;; REFERENCE/DOCKET NUMBER: 27373/8235  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 312/346-5750  
;; TELEFAX: 312/984-9740  
;; TELEX: 25-3856  
;; INFORMATION FOR SEQ ID NO: 3:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 355 amino acids  
;; TYPE: AMINO ACID  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; PCT-US91-06532-3

## alignment\_scores:

Quality: 118.50 Length: 222  
Ratio: 1.411 Gaps: 9  
Percent Similarity: 37.838 Percent Identity: 26.126  
alignment\_block:  
US-09-462-480-2/rev x PCT-US91-06532-3 ..  
Align seg 1/1 to: PCT-US91-06532-3 from: 1 to: 355

```
440 ACGGAGCT.....CACCAGTCCTCTCTCGTCCAGTCGTCCTC 400
|||||
19 ThrGlyAlaValProThrAlaGlnSerGlnValThrSerThrProAsnSe 35
|||||
399 GTCGTCCTTCACGCTCCTCGCGGAGCGGTGGCGCGCGCAGCAGACCG 350
|||||
35 rGluProAlaValArgSerAlaProAlaAlaProProProProA 52
|||||
349 GGTGTGTGGAGCGCGGATTGC..... 327
|||
52 laGlyGlyProProProSerCysSerLeuLeuLeuArgGlnTrpLeuHis 68
.....
326 .....GACCCCTGGCCCATCGC 310
|||
69 ValProGluSerAlaSerAspAspAspAspTrpProAspSe 85
|||
309 TCCC.....GGACCCACCGGAGCGCG.... 288
|||
85 rProProGluSerAlaProGluAlaArgProThrAlaAlaProA 102
|||||
287 ..CCACCCGTACCGACGATCCGGCA.....ACA 261
|||||
102 rgProGlyProHisArgProAlaTrpAlaArgGlyAlaGlyLeuThr 118
|||
260 GCCGCCGCCATCACCGAGGGGCAACCGGCTTTTCGATCAGCTGAGACAT 211
|||||
119 .ProProThrProProArgAlaProSerAlaPheArgAlaSerProS 135
|||||
210 CAGCGGGTGGGTCAACGACCCACCTGCGCCAGGTAGCGACTCCCGCG 161
|||||
135 erAlaCysAlaSerProArgSerThrTrpArgAlaCysAlaCysAspAla 151
|||
160 GCAGCAGGC..... 152
|||||
152 ArgAlaGlyArgGlyArgSerProProArgProProArgProProAr 168
|||
151 .....CCGCGCGCGCTGGGCTGATCCACGACGACGCGG 115
|||||
168 gProProArgProProArgProProArgGlyCysAlaSerArgProThrS 185
|||
114 ATGTTTCGACAGCGGACTGTGGCGGACGAGGCCCATCTGCG..... 74
|||||
185 erGlyCysAlaThrTrpTrpSerGlyProArgProProAlaTrpArgAla 201
|||
74 ..... 74
.....
202 AlaAlaArgGlyProAlaSerGlyProThrGlyLeuGlySerGlyAlaGl 218
|||
73 ....CGGCTTCCTCGCTGGGTGGTTG.....CCGCGCGCGG 41
|||
218 yTrpArgArgProArgArgSerSerGlyArgAlaTrpGlyProArgProV 235
|||
40 TCGCGCCCACTGG 27
|||
235 alProGlyProTrp 239
```

**THIS PAGE BLANK (USPTO)**

OM of: US-09-462-480-2 to: Pending\_Patents\_AA\_Main:\* out\_format : pfs  
Date: Jul 22, 2002 1:35 AM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:  
-MODEL=framel\_n2p\_model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US09462480/runat\_18072002\_164419\_19468/app\_query.fasta\_1.2850  
-DB=Pending\_Patents\_AA\_Main -OFMT=fastan -SUFFIX=rapm  
-GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -XOAPCL=0.000  
-LOOPEXT=0.000 -XGAPOP=4.500 -XGAPEXT=0.050 -XGAPOP=10.000  
-XGAPEXT=0.500 -XGAPOP=6.000 -XGAPEXT=7.000 -XGAPOP=10.000  
-XGAPEXT=0.500 -XGAPOP=6.000 -XGAPEXT=7.000 -XGAPOP=10.000  
-MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOALIGN=200  
-THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pfs -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09462480 -CGN1\_1\_271 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

## Search information block:

Query: US-09-462-480-2  
Query length: 524  
Database: Pending\_Patents\_AA\_Main:\*  
Database sequences: 3502263  
Database length: 351980561  
Search time (sec): 542.380000

## score\_list:

Sequence	Strd Orig	ZScore	Escore	Len	Documentation
/cgn2_6/ptodata/2/paa/PCTUS_COMB.ppt:US99-03265-109	+	713.00	855.38	1.5e-39	368
/cgn2_6/ptodata/2/paa/PCTUS_COMB.ppt:US99-03268-114	+	713.00	855.38	1.5e-39	368
/cgn2_6/ptodata/2/paa/US086_COMB.ppt:US-08-658-800-109	+	713.00	855.38	1.5e-39	368
/cgn2_6/ptodata/2/paa/US086_COMB.ppt:US-08-659-683-114	+	713.00	855.38	1.5e-39	368
/cgn2_6/ptodata/2/paa/US086_COMB.ppt:US-08-680-573-109	+	713.00	855.38	1.5e-39	368
/cgn2_6/ptodata/2/paa/US086_COMB.ppt:US-08-680-574-114	+	713.00	855.38	1.5e-39	368
/cgn2_6/ptodata/2/paa/US087_COMB.ppt:US-08-729-622-109	+	713.00	855.38	1.5e-39	368
/cgn2_6/ptodata/2/paa/US087_COMB.ppt:US-08-730-510-114	+	713.00	855.38	1.5e-39	368
/cgn2_6/ptodata/2/paa/US089_COMB.ppt:US-08-942-341-109	+	713.00	855.38	1.5e-39	368
/cgn2_6/ptodata/2/paa/US089_COMB.ppt:US-08-942-578-114	+	713.00	855.38	1.5e-39	368
/cgn2_6/ptodata/2/paa/US090_COMB.ppt:US-09-024-753-109	+	713.00	855.38	1.5e-39	368
/cgn2_6/ptodata/2/paa/US090_COMB.ppt:US-09-025-197-114	+	713.00	855.38	1.5e-39	368
/cgn2_6/ptodata/2/paa/US090_COMB.ppt:US-09-072-596-109	+	713.00	855.38	1.5e-39	368
/cgn2_6/ptodata/2/paa/US090_COMB.ppt:US-09-072-967-114	+	713.00	855.38	1.5e-39	368
/cgn2_6/ptodata/2/paa/US091_COMB.ppt:US-09-724-685-114	+	713.00	855.38	1.5e-39	368
/cgn2_6/ptodata/2/paa/US100_COMB.ppt:US-10-084-843-114	+	713.00	855.38	1.5e-39	368
/cgn2_6/ptodata/2/paa/US092_COMB.ppt:US-09-050-739-92	+	713.00	855.31	1.5e-39	371
/cgn2_6/ptodata/2/paa/US092_COMB.ppt:US-09-246-191-92	+	713.00	855.31	1.5e-39	371
/cgn2_6/ptodata/2/paa/US097_COMB.ppt:US-09-791-171-92	+	713.00	855.31	1.5e-39	371
/cgn2_6/ptodata/2/paa/US098_COMB.ppt:US-09-804-980-92	+	713.00	855.31	1.5e-39	371
/cgn2_6/ptodata/2/paa/US099_COMB.ppt:US-09-943-443-92	+	713.00	855.31	1.5e-39	371
/cgn2_6/ptodata/2/paa/US090_COMB.ppt:US-09-050-739-90	-	373.00	455.73	7.1e-17	139
/cgn2_6/ptodata/2/paa/US092_COMB.ppt:US-09-246-191-90	-	373.00	455.73	7.1e-17	139
/cgn2_6/ptodata/2/paa/US097_COMB.ppt:US-09-791-171-90	-	373.00	455.73	7.1e-17	139
/cgn2_6/ptodata/2/paa/US098_COMB.ppt:US-09-804-980-90	-	373.00	455.73	7.1e-17	139
/cgn2_6/ptodata/2/paa/US099_COMB.ppt:US-09-943-443-90	-	373.00	455.73	7.1e-17	139
/cgn2_6/ptodata/2/paa/US092_COMB.ppt:US-09-252-991A-29050	-	155.00	186.08	0.0270	3
/cgn2_6/ptodata/2/paa/US092_COMB.ppt:US-09-252-991A-21297	-	148.00	175.81	0.0803	4
/cgn2_6/ptodata/2/paa/PCTUS_COMB.ppt:US00-30037-144	-	147.50	178.68	0.0848	314
/cgn2_6/ptodata/2/paa/PCTUS_COMB.ppt:US01-11988-592	-	147.50	178.68	0.0848	314
/cgn2_6/ptodata/2/paa/PCTUS_COMB.ppt:US02-08278-1613	-	147.50	178.68	0.0848	31
/cgn2_6/ptodata/2/paa/US098_COMB.ppt:US-09-833-245-592	-	147.50	178.68	0.0848	314
/cgn2_6/ptodata/2/paa/US092_COMB.ppt:US-09-252-991A-17143	+	146.50	175.45	0.1002	4
/cgn2_6/ptodata/2/paa/PCTUS_COMB.ppt:US00-05989-700	-	145.50	179.98	0.1127	200
/cgn2_6/ptodata/2/paa/US099_COMB.ppt:US-09-925-297-700	-	145.50	179.98	0.1127	200
/cgn2_6/ptodata/2/paa/US095_COMB.ppt:US-09-599-287A-2	-	145.50	172.41	0.1182	503
/cgn2_6/ptodata/2/paa/US100_COMB.ppt:US-10-078-547-2	-	145.50	172.41	0.1182	503
/cgn2_6/ptodata/2/paa/US092_COMB.ppt:US-09-252-991A-33033	-	144.50	168.29	0.1405	7
/cgn2_6/ptodata/2/paa/US092_COMB.ppt:US-09-252-991A-21881	-	144.00	167.87	0.1517	7
/cgn2_6/ptodata/2/paa/US092_COMB.ppt:US-09-252-991A-31950	-	143.50	178.58	0.1522	7

/cgn2\_6/ptodata/2/paa/PCTUS\_COMB.ppt:US01-04098A-3487 - 141.50 161.26 0.2278  
/cgn2\_6/ptodata/2/paa/US092\_COMB.ppt:US-09-252-991A-30739 - 140.50 164.68 0.2579  
/cgn2\_6/ptodata/2/paa/US092\_COMB.ppt:US-09-252-991A-24973 - 140.00 167.58 0.2724  
/cgn2\_6/ptodata/2/paa/US095\_COMB.ppt:US-09-599-287A-24 - 140.00 165.75 0.2756  
/cgn2\_6/ptodata/2/paa/US100\_COMB.ppt:US-10-078-547-24 - 140.00 165.75 0.2756

seq\_name: /cgn2\_6/ptodata/2/paa/PCTUS\_COMB.ppt:US99-03265-109

seq documentation block:  
; Sequence 109, Application PC/TUS9903265  
; GENERAL INFORMATION:  
; APPLICANT: Corixa Corporation  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF  
; TITLE OF INVENTION: TUBERCULOSIS  
; NUMBER OF SEQUENCES: 350  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds, LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036-2811  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US99/03265  
; FILING DATE: 17-FEB-1999  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 09/024,753  
; FILING DATE: 18-FEB-1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Laura A. Coruzzi  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 9532-0023-228  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-493-4935  
; TELEFAX: 650-493-5556  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 109:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 368 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; PCT-US99-03265-109

alignment\_scores:  
Quality: 713.00 Length: 143  
Ratio: 5.057 Gaps: 0  
Percent Similarity: 98.601 Percent Identity: 96.503  
alignment\_block:  
US-09-462-480-2 x PCT-US99-03265-109 ..  
Align seg 1/1 to: PCT-US99-03265-109 from: 1 to: 368  
1 CTGACGAGGTGACGTGTTGTTACGACGAGTGGCGGCACCGCGCGG 50  
|||||  
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGly 242  
51 CAACCCAGCCGACGAGGAGCGCGGCACATGGCTGCTGGCAGCAGTC 100  
|||||  
242 YAsnProAlaAspGluGluAlaAlaGlnMetGlyLeuGlyThrSerP 259  
|||||  
101 CGCTCTGCAACCATCCGCTGGCTGGTGGATGAGCCAGCCAGCGCGCG 150  
|||||  
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275



```
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 109:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 368 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-658-800-109

alignment_scores:
  Quality: 713.00      Length: 143
  Ratio: 5.057        Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503

alignment_block:
US-09-462-480-2 x US-08-658-800-109 ..
Align seg 1/1 to: US-08-658-800-109 from: 1 to: 368
1 CTCAGCAGGTGACGTGTTTCAGCCAGGTGGCGGCACCGCGCGG 50
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGly 242
51 CAACCCAGCGCAGGAGAACCGCGCAGATGGCCCTGCTCGGCACCAAGTC 100
242 YAsnProAlaAspGluAlaGlnMetGlyLeuLeuGlyThrSerP 259
101 CGGTGTCGAACCATCCGCTGGCTGGTGATCAGGCCCGCGCGGCGG 150
259 roLeuSerAsnHisProLeuAlaGlySerGlyProSerAlaGlyAla 275
151 GGCCTGCTGCGCGGAGTGCTACCTACCTGCGCAGGTGGCTGCTCGG 200
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292
201 CAGCCCGCTGATGCTCAGCTGATCGAAAGCCGGTTGCCCGCTCGGTGA 250
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309
251 TGCCGGCGGCTGTTCCGGATCGCTGCGTGGCGGCGGCGGCGGCGG 300
309 etProAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
301 GGTCCGGAGCGATGGCCAGGGTTCGCAATCCGGCGGCTCCACCGCC 350
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
351 GGGTCTGCTCGCGCGCACCGCTCGCGCAGGAGCGGTGAAGAACGACG 400
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAspG 359
401 AGGACGACTGGGACGAAGAGCAGCTGG 429
359 luAspAspTrpAspGluGluAspAspTrp 368

seq_name: /cgn2_6/ptodata/2/paa/US086_COMB pep: US-08-659-683-114

seq_documentation_block:
; Sequence 114, Application US/08659683
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; NUMBER OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 128
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
```

seq\_name: /cgn2\_6/ptodata/2/paa/US086\_COMB.pep.US-08-680-573-109

seq\_documentation\_block:  
; Sequence 109, Application US/08680573  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Campos-Neto, Antonio  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF  
; TUBERCULOSIS  
; NUMBER OF SEQUENCES: 128  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED AND BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/680,573  
; FILING DATE: 12-JUL-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.417C4  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 109:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 368 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-680-573-109

alignment\_scores:  
Quality: 713.00 Length: 143  
Ratio: 5.057 Gaps: 0  
Percent Similarity: 98.601 Percent Identity: 96.503

alignment\_block:  
US-09-462-480-2 x US-08-680-573-109

Align seg 1/1 to: US-08-680-573-109 from: 1 to: 368

```
1 CTGCAGCAGGTGACGTGTTTCAGCGAGTGGCGCGCCGCGGG 50
|||||
226 LeuGlnValThrSerLeuPheSerGlnValGlyThrGlyGly 242
|||||
51 CAACCCAGCCGACGAGGAAGCGCGCAGATGGCGCTCGGCACCAATC 100
|||||
242 YAsnProAlaAspGluAlaAlaGlnMetGlyLeuGlyThrSerP 259
|||||
101 CGCTGTGCAACCATCCGCTGGCTGGTGGATCAGGCCCGCGCGCG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlySerGlyProSerAlaGlyAla 275
|||||
151 GGCTGTGTCGCGCGAGTGCCTACCTGCGCGAGTGGTGGTGGCCG 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlySerLeuThrAr 292
|||||
201 CACGCCGCTGATGCTCAGCTGATCGAAAGCCGGTTGCCCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309
|||||
```

```
251 TGCCGGCGGCTGTTCCGCGATCGTCCGTGACGGTGGCGCCGCTCCGGTG 300
|||||
309 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
|||||
301 GGTCCGGGAGCGATGGCGCAGGTTTCGCAATCCGGCGGCTCCACAGCCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
|||||
351 GGGTGTGTCGCGCGCGCACGCTCGCGCAGGAGCGTGAAGAGACGAGC 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspSpG 359
|||||
401 AGGACGACTGGGACGAGGAGGACTGG 429
|||||
359 luAspAspTirAspGluGluAspTrp 368
|||||
```

seq\_name: /cgn2\_6/ptodata/2/paa/US086\_COMB.pep.US-08-680-574-114

seq\_documentation\_block:  
; Sequence 114, Application US/08680574  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Campos-Neto, Antonio  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY  
; TUBERCULOSIS  
; NUMBER OF SEQUENCES: 133  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED AND BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/680,574  
; FILING DATE: 12-JUL-1996  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.411C4  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 114:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 368 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-680-574-114

alignment\_scores:  
Quality: 713.00 Length: 143  
Ratio: 5.057 Gaps: 0  
Percent Similarity: 98.601 Percent Identity: 96.503

alignment\_block:  
US-09-462-480-2 x US-08-680-574-114

Align seg 1/1 to: US-08-680-574-114 from: 1 to: 368

```
1 CTCGACGAGTGCAGTGTGTTTCAGCGAGTGGCGCGCCGCGGG 50
|||||
226 LeuGlnValThrSerLeuPheSerGlnValGlyThrGlyGly 242
|||||
```



```
51 CAACCCAGCCGACGAGGAGCGCGCAGATGGCGCTGCTCGGCACCACTC 100
|||||
242 yAsnProAlaAspGluGluAlaAlaGlnMetGlyLeuLeuGlyThrSerP 259
|||||
101 CGCTGTGCAACCATCCGCTGGCTGGTGGATCAGGCCCGCGGGCGCG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275
|||||
151 GGCCTGTGTCGCGCGGAGTCTACCTGGCGCAGGTGGTGGTGGTGGTGG 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292
|||||
201 CACGCCGCTGATGTCTCAGCTGATCGAAAGCCGTTGCCCGCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309
|||||
251 TGCCGGGGCGGTGGTGGCGGATCGTGGTGGCGGCGCGCGTGGTGGTGG 300
|||||
309 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
|||||
301 GGTCCGGGAGCGATGGCGGAGGTTCGCAATCCGCGCGCTCCACCGCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
|||||
351 GGGTCTGTGTCGCGCGGCGCACCGCTCGCGCAGGAGCGTGAAGAGACG 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspSpG 359
|||||
401 AGGACGACTGGGACGAGAGACGACTGG 429
359 luAspAspTrpAspGluGluAspTrp 368
```

seq\_name: /cgn2\_6/ptodata/2/paa/us087\_COMB.pap:us-08-729-622-109

```
seq_documentation_block:
; Sequence 109, Application US/08729622
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, David R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TITLE OF INVENTION: TUBERCULOSIS
; NUMBER OF SEQUENCES: 132
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08729,622
; FILING DATE: 11-OCT-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 109:
; SEQUENCE CHARACTERISTICS:
```

```
; LENGTH: 368 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-729-622-109

alignment_scores:
  Quality: 713.00      Length: 143
  Ratio: 5.057        Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503

alignment_block:
US-09-462-480-2 x US-08-729-622-109 ..
Align seg 1/1 to: US-08-729-622-109 from: 1 to: 368

1 CTGACGAGGTGACGTGCTTTTCAGCCAGTGGCGGCACCGGGCGGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGly 242
|||||
51 CAACCCAGCCGACGAGGAGCGCGCAGATGGCGCTGCTCGGCACCACTC 100
|||||
242 yAsnProAlaAspGluGluAlaAlaGlnMetGlyLeuLeuGlyThrSerP 259
|||||
101 CGCTGTGCAACCATCCGCTGGCTGGTGGATCAGGCCCGCGGGCGCG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275
|||||
151 GGCCTGTGTCGCGCGGAGTCTACCTGGCGCAGGTGGTGGTGGTGGTGG 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292
|||||
201 CACGCCGCTGATGTCTCAGCTGATCGAAAGCCGTTGCCCGCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309
|||||
251 TGCCGGGGCGGTGGTGGCGGATCGTGGTGGCGGCGCGTGGTGGTGG 300
|||||
309 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
|||||
301 GGTCCGGGAGCGATGGCGGAGGTTCGCAATCCGCGCGCTCCACCGCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
|||||
351 GGGTCTGTGTCGCGCGGCGCACCGCTCGCGCAGGAGCGTGAAGAGACG 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspSpG 359
|||||
401 AGGACGACTGGGACGAGAGACGACTGG 429
359 luAspAspTrpAspGluGluAspTrp 368

seq_name: /cgn2_6/ptodata/2/paa/us087_COMB.pap:us-08-730-510-114

seq_documentation_block:
; Sequence 114, Application US/08730510
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 137
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
```

```
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/730,510
; FILING DATE: 27-AUG-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 114:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 368 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-730-510-114
```

```
alignment_scores:
  Quality: 713.00      Length: 143
  Ratio: 5.057         Gaps: 0
  Percent Similarity: 98.601      Percent Identity: 96.503
```

```
alignment_block:
US-09-462-480-2 x US-08-730-510-114
..
Align seg 1/1 to: US-08-730-510-114 from: 1 to: 368
```

```
1 CTGCAGCAGGTGACGTCGTTCTTCAGCCAGGTGGCGGCACCGCGCGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGly 242
51 CAACCCAGCCAGCAGGAGCGGCGCAGATGGCTGCTCGGCACCATC 100
|||||
242 YAsnProAlaAspGluGluAlaGlnMetGlyLeuLeuGlyThrSer 259
101 CGCTGTGGAACATCCGCTGCTGCTGATCAGCCAGCCAGCGCGCGG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGly 275
151 GGCCTGTGCGCGGAGTGCCTACCTGGCGCAGGTGGTGGTTCGACCC 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuTh 292
201 CACGCCGCTGATCTCTAGCTGATCGAAGACCGGTTGCCCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerVal 309
251 TGCCGCGCGCTGTTGCCGATGCTGCTGCTGCTGCTGCTGCTGCTG 300
|||||
309 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaPro 325
301 GGTCCGGAGCGATGGCGGAGGTTCGCAATCCGCGCGCTCCACAGCCC 350
|||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArg 342
351 GGTCTGTGTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAsp 359
401 AGGACGACTGGCGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 429
|||||
359 luAspAspTrpAspGluGluAspAspTrp 368
```

```
seq_name: /cgn2_6/ptodata/2/paa/US089_COMB pep:US-08-942-341-109
seq_documentation_block:
; Sequence 109, Application US/08942341
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedrick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 209
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/942,341
; FILING DATE: 01-OCT-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C7
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 109:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 368 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-942-341-109
```

```
alignment_scores:
  Quality: 713.00      Length: 143
  Ratio: 5.057         Gaps: 0
  Percent Similarity: 98.601      Percent Identity: 96.503
```

```
alignment_block:
US-09-462-480-2 x US-08-942-341-109
..
Align seg 1/1 to: US-08-942-341-109 from: 1 to: 368
```

```
1 CTGCAGCAGGTGACGTCGTTCTTCAGCCAGGTGGCGGCACCGCGCGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGly 242
51 CAACCCAGCCAGCAGGAGCGGCGCAGATGGCTGCTCGGCACCATC 100
|||||
242 YAsnProAlaAspGluGluAlaGlnMetGlyLeuLeuGlyThrSer 259
101 CGCTGTGGAACATCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 150
|||||
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGly 275
151 GGCCTGTGCGCGGAGTGCCTACCTGGCGCAGGTGGTGGTTCGACCC 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuTh 292
```

201 CACCCCTGATGCTCAGCTGATCGAAGGCGGTTGCCCTCGGTGA 250  
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309  
251 TGGCGGGCGGTGTCGGGATCGTCGGTGACGGGTGGCGCGCTCCGGTG 300  
309 etProAlaAlaAlaMetGlyGlnGlySerAlaThrGlyGlyAlaAlaProVal 325  
301 GGTCCGGAGCGATGGCGGCGGTCGCAATCCGGCGGCTCCACGACCC 350  
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342  
351 GGGTCTGGTCCGCGCGCACCGCTCGCGCAGGAGCGTGAAGACGACG 400  
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspG 359  
401 AGGACGACTGGGACGAAGACGACTGG 429  
359 luAspAspTrpAspGluGluAspAspTrp 368

seq\_name: /cgn2\_6/ptodata/2/paa/US089\_COMB.pep:US-08-942-578-114

seq\_documentation\_block:  
; Sequence 114, Application US/08942578  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Campos-Neto, Antonio  
; APPLICANT: Houghton, Raymond  
; APPLICANT: Vedvick, Thomas S.  
; APPLICANT: Twardzik, Daniel R.  
; APPLICANT: Lodes, Michael J.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY  
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS  
; NUMBER OF SEQUENCES: 214  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED AND BERRY LLP  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION NUMBER: US/08/942,578  
; FILING DATE: 01-OCT-1997  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.411C7  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 114:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 368 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-942-578-114

alignment\_scores:  
Quality: 713.00 Length: 143  
Ratio: 5.057 Gaps: 0  
Percent Similarity: 98.601 Percent Identity: 96.503

alignment\_block:  
US-09-462-480-2 x US-08-942-578-114 ..  
Align seq 1/1 to: US-08-942-578-114 from: 1 to: 368  
1 CTCAGCAGAGTGAGCTGCTGTTGTCAGCCAGGTGGCGGCGCACCGCGCGCG 50  
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGlyG 242  
51 CACCCAGCGCAGGAGGAGCCGCGCAGATGGGCTCTCGGCACCAAGTC 100  
242 yAsnProAlaAspGluGluAlaAlaGlnMetGlyLeuLeuGlyThrSerP 259  
101 CGTGTGTCGAACCATCCGCTGGCTGGTGGATCAGGCCAGCCAGCGCGCGCG 150  
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275  
151 GGCTGTGTCGGCGGAGTCTACCTACCTGGCGCAGGTGGGTGGTGGACCCG 200  
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292  
201 CAGCCCTGATGCTCAGCTGATCGAAGGCGGTTGCCCTCGGTGA 250  
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309  
251 TGGCGGGCGGTGTTGCCGGATCGTCGGTGACGGGTGGCGCGCTCCGGTG 300  
309 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325  
301 GGTCCGGAGCGATGGCGGCGGTCGCAATCCGGCGGCTCCACGACCC 350  
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342  
351 GGGTCTGGTCCGCGCGCACCGCTCGCGCAGGAGCGTGAAGACGACG 400  
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspG 359  
401 AGGACGACTGGGACGAAGACGACTGG 429  
359 luAspAspTrpAspGluGluAspAspTrp 368

seq\_name: /cgn2\_6/ptodata/2/paa/US090\_COMB.pep:US-09-024-753-109

seq\_documentation\_block:  
; Sequence 109, Application US/09024753  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Campos-Neto, Antonio  
; APPLICANT: Houghton, Raymond  
; APPLICANT: Vedvick, Thomas S.  
; APPLICANT: Twardzik, Daniel R.  
; APPLICANT: Lodes, Michael J.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF  
; CORRESPONDENCE ADDRESS:  
; NUMBER OF SEQUENCES: 236  
; ADDRESSEE: SEED AND BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/024,753  
; FILING DATE: 18-FEB-1998  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:

TUBERCULOSIS

```
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C8
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 109:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 368 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-024-753-109

alignment_scores:
    Quality: 713.00      Length: 143
    Ratio: 5.057        Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503

alignment_block:
US-09-462-480-2 x US-09-024-753-109 ..
Align seg 1/1 to: US-09-024-753-109 from: 1 to: 368

1 CTGCAGCAGGTGACGTCTGTTTCAGCCAGGTGGCGGCACCGCGCGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGly 242

51 CAACCCAGCCGACGAGGAGCGCGCAGATGGCGCTGCTCGGCACCATC 100
|||||
242 YAsnProAlaAspGluGluAlaGlnMetGlyLeuLeuGlyThrSerP 259

101 CGCTGTGCAACATCCGCTGGTGGTATCAGGCCCGCAGCGCGGCGG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275

151 GGCTGTGTCGGCGGAGTCGCTACCTGGCGCAGGTGGTGGTGGTGG 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292

201 CACCGCGCTGATGCTCAGCTGATCGAAAGCCGCTTCCCGCTCGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValm 309

251 TGCCGGCGGTGTGTCGGGATCGCTGAGCGGTGGCGCGCTCCGCTG 300
|||||
309 etProAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325

301 GGTCCGGAGCGATGGCCAGGTTTCGAATCCGGCGGCTCCACCGCCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342

351 GGGTCTGGTCCGCGCGCAGCCGCTCGCGCAGGACCGCTGAAGAAGC 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluAspAspG 359

401 AGGACGACTGGGACGAGAGGACGACTGG 429
|||||
359 luAspAspTrpAspGluGluAspAspTrp 368

seq_name: /cgn2_6/ptodata/2/paa/US090_COMB.pep:US-09-025-197-114
seq_documentation_block:
; Sequence 114, Application US/09025197
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
```

```
351 GGGTCTGTCGCGCGCACCGCTCGCGCAGGAGCGTGAAGAAGACGACG 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspSpG 359
|||||
401 AGGACGACTGGGACGAGAGACGACTGG 429
|||||
359 luAspAspTrpAspGluGluAspAspTrp 368
|||||
```

seq\_name: /cqn2\_6/ptodata/2/paa/US090\_COMB.pep:US-09-072-596-109

```
seq_documentation_block:
; Sequence 109, Application US/09072596
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 109:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 368 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-072-596-109
```

alignment\_scores:

Quality:	713.00	Length:	143
Ratio:	5.057	Gaps:	0
Percent Similarity:	98.601	Percent Identity:	96.503

alignment\_block:

US-09-462-480-2 x US-09-072-596-109 ..

Align seg 1/1 to: US-09-072-596-109 from: 1 to: 368

```
1 CTGACGAGGTGACGTGCTGTTTCAGCAGGTGGCGGCACCGCGCGGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGly 242
|||||
51 CAACCCAGCCGACGAGGAGCGCGCAGATGGCGCTCTCGGCACCACTC 100
|||||
```

```
242 yAsnProAlaAspGluGluAlaAlaGlnMetGlyLeuLeuGlyThrSerP 259
|||||
101 CGTGTCTGAACCATCCGCTGCTGGTGCATCAGGCCCGCCAGCGGGCGCG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275
|||||
151 GGCTGCTGCGCGCGGAGTGCCTACCTGGCGCAGGTGGTGGTGGTGGTGG 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292
|||||
201 CAGCCCGCTGATGTCTCAGCTGATCGAAAGCCGCGTGGCCCGCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309
|||||
251 TGC CGCGCGCTGTTGCCGATCGCTCGTGACGGGTGGCGGCTCCCGGTG 300
|||||
309 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
|||||
301 GGTCCGCGGAGCGATGGCGCCAGGGTTCCGAATCCCGCGGCTCCACGACC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
|||||
351 GGTGCTGCTGCGCGCGCGCACCGCTCGCGCAGGAGCGGTGAAGAAGACG 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspSpG 359
|||||
401 AGGACGACTGGGACGAGAGACGACTGG 429
|||||
359 luAspAspTrpAspGluGluAspAspTrp 368
|||||
seq_name: /cqn2_6/ptodata/2/paa/US090_COMB.pep:US-09-072-967-114

seq_documentation_block:
; Sequence 114, Application US/09072967
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 114:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 368 amino acids
; 
```

```
;
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-072-967-114

alignment_scores:
  Quality: 713.00      Length: 143
  Ratio: 5.057         Gaps: 0
  Percent Similarity: 98.601  Percent Identity: 96.503

alignment_block:
US-09-462-480-2 x US-09-072-967-114 ..
Align seg 1/1 to: US-09-072-967-114 from: 1 to: 368

1 CTGCAGCAGGTGACGTCGTTGTTTCAGCCAGGTGGCGGCACCGCGCGG 50
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGly 242
51 CAACCCAGCCGACGAGGAGCGCGCAGATGGGCTGCTCGGCACCGATC 100
242 YAsnProAlaAspGluGluAlaGlnMetGlyLeuLeuGlyThrSerP 259
101 CGCTGTGAAACCATCCGCTGGCTGGTGTGATCAGGCCCGCGCGGCG 150
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGly 275
151 GGCCTGTGCGCGGAGTCCCTACCTGCGCGCAGGTGGTGTGACCCG 200
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeu 292
201 CACGCCGCTGATGCTCAGCTGATCGAAAGCCGGTTGCCCGCTCGGT 250
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSer 309
251 TGCCGGCGGCTGTTCGCGATGTCGCGTACGTCGCGGTGGCGCGCT 300
309 etProAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaPro 325
301 GGTCCGGGAGCGATGGCGAGGTTCCGATCCGCGCGCTCCACGACCC 350
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThr 342
351 GGGTCTGTGTCGCGCGCACCGCTCGGCGAGGAGCGGTGAAGAAG 400
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluAsp 359
401 AGGACGACTGGGACGAGAGGACGACTGG 429
359 luAspAspTrpAspGluGluAspAspTrp 368

seq_name: /cgn2_6/ptodata/2/paa/US097_COMB.pep:US-09-724-685-114

seq_documentation_block:
; Sequence 114, Application US/09724685
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
;           Skeiky, Yasir
;           Dillon, Davin C.
;           Campos-Neto, Antonio
; TITLE OF INVENTION: Compounds and Methods for
; Immunotherapy and Diagnosis of Tuberculosis
; NUMBER OF SEQUENCES: 155
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```

309 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
301 GGTCGGGAGCGATGGCCAGGTTCCGAATCCGGCGGCTCCACCGAGCC 350
   ||| ||||| ||||| ||||| ::||| ||||| ||||| |||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
351 GGCTCTGTCGCGCGCCACCGCTCGCGCAGGAGCGGTGAAGAAGACGACG 400
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAspG 359
401 AGGACGACTGGGACGAGAGAGGACTGG 429
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
359 luAspAspTrpAspGluGluAspTrp 368

```

**THIS PAGE BLANK (USPTO)**



Sequence	Strid Orig	zScore	EScore	Len	Documentation
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.p	US-10-080-1708-13	+ 198.00	236.38	6.9e-05	
/cgn2_6/ptodata/2/paa/US09_NEW_COMB.p	US-09-994-404-252	+ 133.00	136.29	0.7130	
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.p	US-10-155-881-220336	+ 127.50	148.90	1.165	
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.p	US-10-155-881-220336	+ 126.50	146.28	1.190	
/cgn2_6/ptodata/2/paa/US60_NEW_COMB.p	US-60-382-898-278	+ 125.50	139.06	2.15	
/cgn2_6/ptodata/2/paa/US09_NEW_COMB.p	US-09-667-170A-187	+ 133.50	138.16	2.88	
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.p	US-10-080-1708-452	+ 123.00	144.95	3.16	
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.p	US-10-155-881-10704	+ 122.00	140.40	3.61	
/cgn2_6/ptodata/2/paa/PCT_NEW_COMB.p	PCT-US02-09921-1022	+ 121.50	139.35	3.87	
/cgn2_6/ptodata/2/paa/US60_NEW_COMB.p	US-60-360-039-3962	+ 121.50	135.39	3.82	
/cgn2_6/ptodata/2/paa/US09_NEW_COMB.p	US-09-784-984A-8	+ 121.00	138.32	4.15	
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.p	US-10-155-881-11183	+ 121.00	134.49	4.10	
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.p	US-10-155-881-10047	+ 120.50	139.26	4.47	
/cgn2_6/ptodata/2/paa/US09_NEW_COMB.p	US-09-573-655B-274	+ 119.50	138.62	5.17	
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.p	US-10-155-881-7219	+ 119.00	138.19	5.56	
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.p	US-10-155-881-8543	+ 118.50	145.11	6.10	
/cgn2_6/ptodata/2/paa/US09_NEW_COMB.p	US-09-540-209B-10118	+ 118.50	139.96	6.01	
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.p	US-10-155-881-26034	+ 118.50	136.65	5.95	
/cgn2_6/ptodata/2/paa/US60_NEW_COMB.p	US-60-360-039-2275	+ 118.50	132.04	5.87	
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.p	US-10-155-881-9390	+ 118.00	135.82	6.39	
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.p	US-10-155-299-6164	+ 117.00	133.85	7.35	
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.p	US-10-155-881-10185	+ 117.00	132.61	7.32	
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.p	US-10-160-568-2	+ 117.00	127.92	7.22	
/cgn2_6/ptodata/2/paa/PCT_NEW_COMB.p	PCT-US02-16676-9	+ 117.00	125.25	7.17	
/cgn2_6/ptodata/2/paa/US60_NEW_COMB.p	US-60-360-039-19227	+ 116.50	128.80	7.79	
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.p	US-10-155-881-11798	+ 116.00	135.07	8.54	
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.p	US-10-121-988-47	+ 116.00	127.30	8.34	
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.p	US-10-155-881-9768	+ 115.50	135.49	9.19	
/cgn2_6/ptodata/2/paa/PCT_NEW_COMB.p	PCT-US02-13874-3	+ 115.50	125.26	8.92	
/cgn2_6/ptodata/2/paa/US60_NEW_COMB.p	US-60-389-987-218	+ 115.50	122.95	8.86	
/cgn2_6/ptodata/2/paa/US60_NEW_COMB.p	US-60-389-987-237	+ 115.50	122.66	8.85	
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.p	US-10-155-881-7239	+ 115.00	133.27	9.83	
/cgn2_6/ptodata/2/paa/US60_NEW_COMB.p	US-60-360-039-20627	+ 115.00	127.04	9.65	
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.p	US-10-155-881-37314	+ 115.00	126.86	9.64	
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.p	US-10-137-337-699	+ 114.50	132.06	10.54	
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.p	US-10-168-659-13	+ 114.50	122.65	10.25	
/cgn2_6/ptodata/2/paa/US09_NEW_COMB.p	US-09-819-104A-2	+ 114.50	117.16	10.09	
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.p	US-10-138-145-850	+ 114.50	136.61	12.36	
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.p	US-10-104-047-338	+ 113.50	132.78	12.22	
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.p	US-10-138-098-52	+ 113.50	126.44	12.00	

## seq\_documentation\_block:

```
; Sequence 252, Application US/09994404
; GENERAL INFORMATION:
; APPLICANT: ESSER, KLAUS M.
; APPLICANT: CHAN, JOHN Y.
; APPLICANT: DABROWSKI-AMARAL, CHRISTINE ELLEN
; APPLICANT: DELVECCCHIO, ALFRED MICHAEL
; APPLICANT: DILLON, SUSAN B.
; APPLICANT: LEARY, JEFFREY JOSEPH
; APPLICANT: SUTTON, DAVID
; TITLE OF INVENTION: NOVEL CODING SEQUENCES FROM HERPES
; TITLE OF INVENTION: SIMPLEX VIRUS TYPE-2
; FILE REFERENCE: P50583
; CURRENT APPLICATION NUMBER: US/09/994,404
; CURRENT FILING DATE: 2001-11-26
; PRIOR APPLICATION NUMBER: 09/297,477
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 60/030,279
; PRIOR FILING DATE: 1999-11-04
; PRIOR APPLICATION NUMBER: US 60/049,018
; PRIOR FILING DATE: 1997-06-09
; NUMBER OF SEQ ID NOS: 303
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 252
; LENGTH: 3051
; TYPE: PROT
; ORGANISM: Herpes simplex
US-09-994-404-252
```

## alignment\_scores:

```
Quality: 133.00 Length: 253
Ratio: 1.357 Gaps: 13
Percent Similarity: 38.735 Percent Identity: 26.877
```

## alignment\_block:

```
US-09-462-480-2/rev x US-09-994-404-252 ..
```

```
Align seg 1/1 to: US-09-994-404-252 from: 1 to: 3051
```

```
502 CTTCTCTCGCAAAATGTCGAGTCTTCGGCGCGGGTGGCGGGAAG 453
||||| ||| : : : ||| : : : ||| : : : ||| : : : |||
2582 ProSerProAlaAspAlaAlaLeuProProAlaPheSerGlyse 2598
| : : : : : : : : : : : : : : : : : : : : : : :
452 TCTGTCCTCATTCAGGAGTCACAGTCGCTCTTCGTCGCCAGTCGC 403
| : : : : : : : : : : : : : : : : : : : : : : :
2598 rAlaAlaAlaPheSerAlaAlaValProArgValArgSerArgT 2615
402 CTCGTCCTCTCTCAGCTCTCTCGCGGAGGGTGCGG..... 365
||| : : : ||| : : : ||| : : : ||| : : : |||
2615 hrArgAlaLysSerArgAlaProAlaSerAlaProGluGlyTrp 2631
364 ..... : : : : : : : : : : : : : : : : : : : :
2632 ArgProProAlaLeuProAlaProValAlaProValAlaAlaSerAlaAr 2648
359 ACCA..... : : : : : : : : : : : : : : : : : : : :
||| : : : ||| : : : ||| : : : ||| : : : |||
2648 gProProAspGlnProProThrProGluSerAlaProProAlaTrpValS 2665
339 GCGCGCGGATGGAAACCTCGGCCATCG..... : : : : : : : : :
||| : : : ||| : : : ||| : : : ||| : : : |||
2665 erAlaLeuProLeuProGlyProAlaSerArgAlaPheProAlaPro 2681
310 ..... : : : : : : : : : : : : : : : : : : : :
2682 ThrLeuAlaProIleProProProAlaGluGlyAlaValAlaProGI 2698
295 .... GAGCGGGCGGCAC..... : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : :
2698 yAspAspArgArgGlyArgGlnThrThrAlaGlyProSerProT 2715
273 CGATCCGGCAACAGCCCGCGCA..... : : : : : : : : : : : :
r x
```

```
|| : : : ||| |||
2715 hrProProArgGlyProAlaAlaGlyProProArgLeuTrpAlaVal 2731
250 ..... : : : : : : : : : : : : : : : : : : : :
2732 AlaSerLeuSerAlaSerLeuAsnSerLeuProSerProArg...AspPr 2747
233 GGCCTTTTCGATCAGTCAGACATCAGGGCGTGGGGTCAACGACCCACC 184
||| : : : : : : : : : : : : : : : : : : : :
2747 oAlaAspHisAlaAlaAlaValSerAlaAlaAlaAlaVal...ProPr 2763
183 TGCGCCAGGT.....AGCGACTCCGCGCAGCAGAGCCCGCGCCGC 143
| : : : ||| : : : ||| : : : ||| : : : |||
2763 oSerProGlyLeuAlaProProThrSerAlaValGlnThrSerProProp 2780
142 CGCTGGGCGCTGATCCA..... : : : : : : : : : : : : : :
||| : : : ||| : : : ||| : : : ||| : : : |||
2780 roLeuAlaProGlyProValAlaProSerGluProLeuLeuCysGlyTrp... 2795
107 GACAGCGGACTGTGTCGAGCAGCGCCATCTGCGCGGCTTCCTCGTCGCG 58
: : : : : : : : : : : : : : : : : : : :
2796 .....ValValPro.GlyGlyProValAlaArgPro..... 2806
57 TGGTTGCCGCGCGGTGGCGCCACCTGGCTGAACAACAGCGTCACCT 8
||||| : : : : : : : : : : : : : :
2807 .....ProProGlnSerProAlaThrLysPro 2815
7 GCTGCA 2
|||||
2816 AlaAla 2817
seq_name: /cgn2_6/ptodata/2/paa/us10_NEW_COMB.pep:US-10-155-881-22036
seq_documentation_block:
; Sequence 22036, Application US/10155881
; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Lutfiyya, Linda L.
; APPLICANT: McIninch, James
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TITLE OF INVENTION: TRANSCRIPTION IN PLANTS
; FILE REFERENCE: 38-21(15300)J
; CURRENT APPLICATION NUMBER: US/10/155,881
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 37595
; SEQ ID NO 22036
; LENGTH: 261
; TYPE: PROT
; ORGANISM: Glycine max
US-10-155-881-22036
```

## alignment\_scores:

```
Quality: 127.50 Length: 181
Ratio: 1.401 Gaps: 11
Percent Similarity: 50.276 Percent Identity: 35.359
```

## alignment\_block:

```
US-09-462-480-2/rev x US-10-155-881-22036 ..
```

```
Align seg 1/1 to: US-10-155-881-22036 from: 1 to: 261
```

```
476 CTTCCGCGCGCGGTGGCGGAGTCTGTGTCATTACGGGAGCTCACCA 427
||||| : : : : : : : : : : : : : : : : : : : :
65 LeuProArgThrAlaThrAlaArgSerThrAlaAlaGlyAlaAlaPheAl 81
426 GTCGTCCTCTTCGTCACAGTCGTCCTGCTCTTCTTTCAGCTCCCTCG 377
||||| : : : : : : : : : : : : : : : : : : : :
81 aCysArgSerSerAlaProProAlaSerSerSerSerArgAlaSerSera 98
376 CGAGCGGTGCGG...GCGCGACCAAGACCCCGGCTGGTGGAGCGCCGGAT 330
```

```
98 laThrSer.ProThrAlaArgProSerSerGlySerAlaArgArg.. 113
|||||
329 TGCACACCTGCCCATCG...CTCCGGACCCACCGGCGGCCACC 283
|||||
114 ....AsnProProSerProLeuProAlaProAlaProLeuArgArg 129
|||||
282 CGTCACGACGATCCGCAACAGCCGCGGCATCACCAGGGGCAACCG 233
|||||
129 laSerProLeu.....SerProThrThrSerProSerSerLeuPro 142
|||||
232 GCTTTTCATCAGTCAGATCAGCGCGGTGCGGGTCACGACCCACCT 183
|||||
143 LeuProProSerSerAlaAsnAlaPheAlaLeuThrThrThr..Pro 158
|||||
182 GCGCAGGTAGCGACTCCGCGCAGCAGGCC.....GCGCCGC 142
|||||
159 ProProLysThrThrProSerArgSerSerProProProArgArgProPr 175
|||||
141 GCTGGGCGCTGATCCACAGCAGCGATGTTGCGACAGCGGACTGGTGC 92
|||||
175 oSerGlyProTyrLeuProAlaProThrSerAlaArgSerGlyAlaSerP 192
|||||
91 CGACGAGGCC...ATTCGCGCGCTTCCTCGTCGGTGGGTGC..... 50
|||||
192 roProArgProAsnSerCys.ArgCysArgArgArgilleGlnCysPheTh 208
|||||
49 .....CSCCGCGGTGCGGCCACCT..... 29
|||||
208 rIleThrIleAsnThrThrSerSerAsnSerAsnArgProProTrpGluA 225
|||||
28 .....GGCTGAACAACGACGTCACCTGTCT 5
|||||
225 rgProLeuProGlnGlySerGlyIleThrSerProAla 237
|||||
```

seq\_name: /cgn2\_5/ptodata/2/paa/US10\_NEW\_COMB.pep:US-10-155-881-25973

```
seq_documentation_block:
; Sequence 25973, Application US/10155881
; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Lutfiyya, Linda L.
; APPLICANT: McIninch, James
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TITLE OF INVENTION: TRANSCRIPTION IN PLANTS
; FILE REFERENCE: 38-21(15300)J
; CURRENT APPLICATION NUMBER: US/10/155,881
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 37595
; SEQ ID NO 25973
; LENGTH: 318
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-155-881-25973
```

alignment\_scores:  
Quality: 126.50 Length: 199  
Ratio: 1.360 Gaps: 12  
Percent Similarity: 46.734 Percent Identity: 30.653

alignment\_block:  
US-09-462-480-2 x US-10-155-881-25973 ..  
Align seg 1/1 to: US-10-155-881-25973 from: 1 to: 318  
..  
12 GACGTCTGTTTCAGCGGCGGCGCACCGCGCGGCGGCAACCCACCG 61  
|||||  
120 AspalapheProArgAlaGlyAspArgGlnArgArgHis.....Ar 134  
|||||  
62 ACGAGGAAGCGCGCAGATGGCGCTGCTCGGCACCGCTGCTCGAAC 111

```
|||||
134 gArgGly.....HisArgGlyLeuLeuPro 143
|||||
112 CATCCGTGGCTGGTATCAGCCCGCAGCGCGCGCGC..... 149
|||||
143 roGlnAlaAlaArgLeuAlaGlnArgGluArgArgHisGln 159
|||||
150 .....GGGCCTGTCGCGCGGAGTACCTACCTGGCGCAGG..... 185
|||||
160 ArgHisAlaProAlaAlaArgGly.....AspTrpGlyArgArgArgAr 174
|||||
186 .....TGGTCTGTTACCGG..... 200
|||||
174 gAlaAlaGlyGluValArgAspIleValHisValTrpArgLeuProG 191
|||||
201 .....CACGCCGCTGATGCTCAGCTGATCGGAAA 230
|||||
191 lYalaGlyAlaAlaArgGlyHisGlyAlaArgValProArgArgArg 207
|||||
231 GCCGTTGCCCTCGGTGATCGCGCGGCTGTTGCCGATCGTCGGTGA 280
|||||
208 AlaGly.....AlaGlyGly..... 212
|||||
281 CGGTGGCGCGCTCCGTCGGGAGCGATGGCCAGGGTTCGCAA 330
|||||
213 .GlyTrpGluArgHisGlyGlyAlaAspArgValGlyProArgHisGly 229
|||||
331 TCCCG.....CGCTCCACAGCCCGGTCTGTCGCGCGCGCAC 371
|||||
229 spArgGlyHisValArgGlnArgHisValArgGluAlaAlaGlyPro 245
|||||
372 GTCGCGCAGGAGCGTGAAGAAGACGACGAGCGAGTGGGACGAAGAG 421
|||||
246 GlyArgArgGlyGly.....ArgArgAlaValArgValGlyArgArg 260
|||||
422 ACGA.....CTGCTAGCTCCCGTAATGACAACAGACTTCCCGGCC 462
|||||
260 gArgAlaAspGlyAlaAlaGluGlnArgArgArgArgAlaProAla 277
|||||
463 ACCCGCGCGGAAGACTTGCACACATTTTGGCGGAGGAAGTAAG 507
|||||
277 spValArgArg.ArgProAlaAspAlaAlaAlaArgHisValArg 291
|||||
```

seq\_name: /cgn2\_6/ptodata/2/paa/US60\_NEW\_COMB.pep:US-60-382-898-278

```
seq_documentation_block:
; Sequence 278, Application US/60382898
; GENERAL INFORMATION:
; APPLICANT: Hudson, Keith
; APPLICANT: et al.
; TITLE OF INVENTION: Plant Receptors and Ligands
; FILE REFERENCE: 1066P
; CURRENT APPLICATION NUMBER: US/60/382,898
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 1344
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 278
; LENGTH: 708
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-60-382-898-278
```

alignment\_scores:  
Quality: 125.50 Length: 175  
Ratio: 1.793 Gaps: 7  
Percent Similarity: 40.000 Percent Identity: 29.143

alignment\_block:  
US-09-462-480-2/rev x US-60-382-898-278 ..  
Align seg 1/1 to: US-60-382-898-278 from: 1 to: 708

```
421 CCTCTCTCCAGTCGTCCTCGTCTTCTTACAGCTCCTCGCGCAGC 372
||||| ||||| ||| |||||
38 ProLeuProProSerAlaPro.....ProProAsnArgAl 49
||||| ||||| |||||
371 GGTGCGCGCGCAGCAGCCCGGGTGTGGAGCGCGCGGATTGCGAAC. 323
| ||| ||| |||||
49 aProProProProProProValThrSerProProValAlaAsnG 66
||||| ||||| ||||| |||||
322 .....CCTGGCCATCGCTCCGGACCCAGCGGAGCGCGGCAC..... 284
||| ||| ||||| ||||| |||||
66 lAlaProProProProLeuProLysProProGluSerSerProPro 82
||||| ||||| ||||| |||||
283 CCGTCACCGAGATCCGCAACAGCCCGCGCA...TCACCGAGGGGCA 237
||||| ||||| ||||| ||||| |||||
83 ProGlnProValIleProSerProProSerThrSerProProProG1 99
||||| ||||| ||||| |||||
236 ACCGGCTTTTCAGTCAGTCAGACATCAGCGCGTG..... 201
||||| ||||| ||||| |||||
99 nProValIleProSerProProProSerAlaSerProProAlaLeuV 116
||||| ||||| ||||| |||||
201 ..... 201
116 alProProLeuProSerSerProProProProAlaSerValProPro 132
||||| ||||| ||||| |||||
201 ..... 201
133 ArgProSerProSerProProIleLeuValArgSerProProProSerVa 149
||||| ||||| ||||| |||||
200 .....CGGTCAACGACCCACCTCGCCAGGTAGCGACTCCGCGCGCAGCA 155
||||| ||||| ||||| |||||
149 largProIleGlnSerProProProProProSerAspArgProThrGlns 166
||||| ||||| ||||| |||||
154 GGCGCGCGCGCGTGGCGGCTGATCCACCGAGCGCGGATGTTTCGAC 105
||||| ||||| ||||| |||||
166 erProProProProSerProProSerProProSerGluArgProThrGln 182
||||| ||||| ||||| |||||
104 AGCGGACTGTGCGGAGAGCGCCATCTGCGCGGCTTCTCTCGCGGTGG 55
||||| ||||| ||||| |||||
183 Ser.....ProProSerProSerGluArgProThrGlnSerPro... 196
||||| ||||| ||||| |||||
54 GTTGGCGCGCGGTGGCGCGCAC 30
||||| ||||| ||||| |||||
197 ....ProProSerProProSer 203
```

seq\_name: /cgn2\_5/ptodata/2/paa/US09\_NEW\_COMB.pep:US-09-667-170A-187

seq\_documentation\_block:

; Sequence 187, Application US/09667170A

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; APPLICANT: Henderson, Robert A.

; APPLICANT: Lodes, Michael J.

; APPLICANT: Fling, Steven P.

; APPLICANT: Mohamath, Radoh

; APPLICANT: Algate, Paul A.

; APPLICANT: Secrist, Heather

; APPLICANT: Indirias, Carol Yoseph

; APPLICANT: Benson, Darin R.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR

; THE THERAPY AND DIAGNOSIS OF LUNG CANCER

; FILE REFERENCE: 210121.475C7

; CURRENT APPLICATION NUMBER: US/09/667,170A

; CURRENT FILING DATE: 2002-05-03

; NUMBER OF SEQ ID NOS: 440

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 187

; LENGTH: 595

; TYPE: PRT

; ORGANISM: Homo sapien

US-09-667-170A-187

alignment\_scores:

% \* . -

Quality: 123.50 Length: 170  
Ratio: 1.816 Gaps: 8  
Percent Similarity: 40.000 Percent Identity: 28.824

alignment\_block:

US-09-462-480-2/rev x US-09-667-170A-187 ..

Align seg 1/1 to: US-09-667-170A-187 from: 1 to: 595

```
490 AAATGTTTGGCAAGTCTCCGG.....CCCGGGTGGCGGGGAAGTCTGTT 447
::: ||||| ||||| |||||
474 ArgLeuTrpArgGlnPheargvalcglyglyPhePro..... 486
446 GTCATTACGGGAGCTCACAGTCGTCCTTCTGTCCTCCAGTCGTCCTCGTC 397
||| ||||| |||||
487 .....ProProProProSerArgProProA 495
396 GTCCTCTTCA.....CGCTCCTGCGGAGCGGTGCCGGCGG 360
::: ||||| ||||| |||||
495 laValLeuLeuProLeuLeuArgLeuAlaCysAlaGlyAspProGlyAla 511
359 ACCAGACCCGGGCTGGTGAGCGCGCGGATTGCGAACCTGCGCCCATCGC 310
||||| ||||| |||||
512 ThrArgProGlyProArgArgPro.....Al 520
309 TCCCGGACCCACCGGAGCGCGCCACCGTCACCGACGATCGGGCAACA. 261
||| ||| ||| ||| |||||
520 aArgArgProArgGlyGluLeuIleProArgProArgProAlaAlap 537
260 .....GCCGCGGCATCACCGAGGGGCAACC 234
537 roSerGluGluGlyLeuArgMetGluSerSerValAspGlyAlaThr 553
233 GGCCTTTTCGATCAGTCAGATCAGCGCGGTGCGGGTCAACGACCCACC 184
553 ..... 553
183 TCGCCAGGTAGCGACTCCGCGCGCAGCAGCCCGCGCGGTGGGCG 134
||| ||||| ||||| |||||
554 .AlaThrThrAlaAspAlaAla...SerGlyGluAlaProGluAlaGlyP 569
133 CTGATCCACCGACCGCGGATGTTTCGACAGCGGACTGTCGCCGAGCAGG 84
||||| ||||| |||||
569 roSerProSerHisSer..... 574
83 CCCATCTCGCGGCTTCTCTCGTCGCTGCGGTGCGCGCGGTGCCGCC 34
||| ||||| ||||| |||||
575 ProThrMetCysGlnThrGlyGlyProGlyProProProGlnProPr 591
33 CACCTGGCTG 24
||| |||||
591 oArgTriLeu 594
```

seq\_name: /cgn2\_5/ptodata/2/paa/US10\_NEW\_COMB.pep:US-10-080-170B-452

seq\_documentation\_block:

; Sequence 452, Application US/10080170B

; GENERAL INFORMATION:

; APPLICANT: COLE, S.T.

; TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR

; IDENTIFYING TARGETS FOR THE DIAGNOSIS, PHYLOGENESIS OR

; TREATMENT OF MYCOBACTERIOSES

; FILE REFERENCE: 03495.0218

; CURRENT APPLICATION NUMBER: US/10/080,170B

; CURRENT FILING DATE: 2002-06-10

; PRIOR APPLICATION NUMBER: 60/270,123

; PRIOR FILING DATE: 2001-02-22

; NUMBER OF SEQ ID NOS: 652

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 452

; LENGTH: 227

; TYPE: PRT

; ORGANISM: Mycobacterium tuberculosis

US-10-080-170B-452

alignment\_scores:  
Quality: 123.00 Length: 220  
Ratio: 1.281 Gaps: 11  
Percent Similarity: 43.636 Percent Identity: 26.364

alignment\_block:  
US-09-462-480-2/rev x US-10-080-170B-452 ..  
Align seg 1/1 to: US-10-080-170B-452 from: 1 to: 227

464 GTGGCCGGGAAGTCTGTGTCTATACGGGA.....GCTCAC..... 429  
|||||.....  
15 ValGlySerSerAlaAlaLeuLeuThrGlyGlyLeAlaHisAlaAspPr 31  
428 .....CAGTCGTCTCTTCGT 413  
31 oAlaProAlaProAlaProAlaProAsnIleProGlnGlnLeuIleSerS 48  
412 CCCAGTCGTCTCGTCGCTTCTTCACGCTCTCTGCGGAGCGGTCCCGC 363  
|||.....  
48 erAlaAlaAsnAlaProGlnIleLeuGlnAsnLeuAlaThrAlaLeuGly 64  
362 GCGACACAGCCCGGCTGGTGGAGCCGCCGATTGCGAACCCCTGCCCAT 313  
||||| ||| ||| ||| .....  
65 Alathr....ProLeuSerAlaProLysValAlaGluProAlaProAl 80  
312 CGTCCCGGACCCACCGGAGCGGCCACCGTCACCGAC...GATCCGG 266  
||||| ||| ||| ||| .....  
80 aAlaProGlyIleThrAlaThrPheProGlyLeuThrProAlaAlaProA 97  
265 CAACAGCCGCC..... 255  
||||| ||| ||| .....  
97 laAlaAlaAlaAlaProAlaLeuThrProSerIleProGlyValAsnAla 113  
254 .....GGCATCACCGAGGGGCAACCGCTTT..... 228  
114 ProIleProGlyIleThrProAlaAlaProAlaLeuProValThrAlaPr 130  
227 .....TCGATCAGCTGACATCAGCGCGCTGC 200  
130 oAlaAlaAlaProThrIleProGlyValAsnAlaProIleProGlyIleT 147  
199 GGGTCAACGACCCACTCGG.....CCAGGTAGC 171  
147 hrAlaProAlaProAlaAlaAlaValProAlaSerValProGlyVal 163  
170 GACTCCGGCGCG..... 159  
164 ProSerAlaLysValAspLeuProGlnLeuProTyrLeuProLeuGlnVa 180  
158 .....ACGAGCGCGCG.....CCGCGCTGGGCGCTGATC 128  
180 lProGlnGlnLeuSerLeuProAlaAspLeuProAlaLeuAla..... 194  
127 CACACGACGAGGATGGTTCGACAGCGGACTGGTCCGCGAGCAGGCCATC 78  
195 .....SerGlyValIleProAlaAlaProIle 203  
77 TCGCGGGCTTCTCTCGTGGCTGGGTGGCG.....CCGCGGTGCCGCC 34  
204 AlaProThrProAlaProGlyAlaProAlaLeuProProGlyProPr 220  
33 CACTGGGTG 24  
||| |||  
220 oSerLeuLeu 223

seq\_name: /cgn2\_6/ptodata/2/paa/US10\_NEW\_COMB.pep:US-10-155-881-10704

seq\_documentation\_block:  
; Sequence 10704, Application US/10155881

GENERAL INFORMATION:

; APPLICANT: Dotson, Stanton B.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Lutfiyya, Linda L.  
; APPLICANT: McIninch, James  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH TRANSCRIPTION IN PLANTS  
; FILE REFERENCE: 38-21(15300)J  
; CURRENT APPLICATION NUMBER: US/10/155,881  
; CURRENT FILING DATE: 2002-05-22  
; NUMBER OF SEQ ID NOS: 37595  
; SEQ ID NO 10704  
; LENGTH: 356  
; TYPE: PRT  
; ORGANISM: Zea mays  
US-10-155-881-10704

alignment\_scores:  
Quality: 122.00 Length: 154  
Ratio: 1.649 Gaps: 9  
Percent Similarity: 48.052 Percent Identity: 34.416

alignment\_block:  
US-09-462-480-2/rev x US-10-155-881-10704 ..  
Align seg 1/1 to: US-10-155-881-10704 from: 1 to: 356

502 CTTCTCTCGCAAAATGTTGCAAGTCTTCGCGCCGGTGGCGGGAAG 453  
||||| ||| ||| .....  
207 ProSerSerSerThrThrThrThrThrThrProAlaAlaPro..... 221  
452 TCTGTTGTCATTACGGGAGCTCACC...AGTCGTCTCTTGTCCAGTC 406  
||| ||| ||| ||| .....  
222 ...LeuProCysArgArgCysThrProProThrProThrArgSerThrA 237  
405 GTCCTCTCGCTCTTCTCACGCTCTCTGGCGA.....GCGGTGCGG 365  
||||| ||| ||| .....  
237 rgGlnArgArgArgGlnArgThrProAlaArgThrAlaAlaAlaSerPro 253  
364 GCGGACACGACCGCGGCTGTGGAGCCGCGGATTCGAAACCCGTGGCCC 315  
:||| ||| ||| .....  
254 ThrArgProProPro.....ArgAlaProProThrSerPr 265  
314 ATCGTCCCGGACCCACCGGAGCGG.....C 289  
||||| ||| ||| .....  
265 oSerLeuProProProProAlaSerThrArgValAlaThrThrTrpProA 282  
288 GCCACCCGTCACCGACATCCGCAACAGCCGCGCATCACCGAGGGG 239  
||| ||| ||| ||| .....  
282 rgAlaProArgArgSerHisArgAlaArgAlaProProSerProArg 298  
238 CAACGCGCTTTTCGATCAGCTGAGACATCAGCGGCGTGGGTCAACGAC 189  
||| ||| ||| ||| .....  
299 ArgArgArg...ArgAlaProThrThrArgSerArgCysProSerSerTh 314  
188 CCACCTGGCGGAGTAGCAGCTCCCGCGCAGCAGGCCCG.....CGCCCGC 142  
| .....  
314 rSerSerAlaSerAlaArgArgAspGlyGlyAlaGlyProGlyArgProG 331  
141 GCTGGGCGCTGATCCACCGACGAGCGGATGGTTTCACAGCGGACTGGTGC 92  
||| ||| ||| ||| .....  
331 lyGlyGlyVal.....GlnGluGlyGlyGlyGlyGlnArgThrArgAl 345  
91 CGAGCAGGCC 82  
| |||  
345 aProAlaAla 348

seq\_name: /cgn2\_6/ptodata/2/paa/PCT\_NEW\_COMB.pep:PCT-US02-09921-1022

seq\_documentation\_block:  
; Sequence 1022, Application PC/TUS0209921

```

329  TCGGAACCTCGGCCATCGTCTCCGG.....GACCCACGGAGCGCGCC 286
    |||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
187  GluGlu.ArgProAlaAlaLeuProSerAlaSerProProArgGluV 203
    285  ACCCGTCACCGAGC.....ATCCGGCAACAGCCG 257
    |||  |||||  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
203  alProValProThrProAlaProValGluValProValProGluArgVal 219
    256  CCGGCATCACCGAGGGGCAACCGCTTTCGATCAGCTGACGACATCAGC 207
    |||  |||||  |||||  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
220  AlaGlySerProValThrProProAlaArgThrGlyGlyValSerSerly 236
    206  GCGGTGC.....GGGTCAACGAGCCACCTGCGCCAGGTAGCGACTC 166
    |||||  |||||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
236  sAlaCysArgProHisGlyGlyValThrProGlnCysAlaSerAlaSerP 253
    165  CGCGGCGCAGCA.....GGCCGGCGCGCGCTGGGGCTGTATCCACCAAG 122
    |||||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
253  roArgThrThrCysTyrAlaProArgPro.....LeuProHisAla 266
    121  CCAGCGGAGTGTTCGACAGCGGACTGTGTGCGGACGAGGCCCATCTCGCGG 72
    |||  |||||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
267  ProMet.....SerValPro..LeuLeuProSer...ProSerCysPro 279
    71  GCTTCTCTCGTGGCTGGG...TTGCGCGCGCGGTGCGCGCCACCTGG 27
    ::|||  ::|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
280  HisProArgAsnAlaGlyLysLeuSerProSerLeuProSerArgTrp 295
    seq_name: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep:US-60-360-039

seq_documentation_block:
; Sequence 3962, Application US/60360039
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Chen, Xianfeng
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)A
; CURRENT APPLICATION NUMBER: US/60/360,039
; CURRENT FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 3962
; LENGTH: 639
; TYPE: PRT
; ORGANISM: Neurospora crassa
; US-60-360-039-1962

```

```

alignment_scores:
  Quality: 121.50      Length: 189
  Ratio: 1.538        Gaps: 12
  Percent Similarity: 41.799    Percent Identity: 31.217

alignment_block:
US-09-462-480-2/rev x US-60-360-039-3962  ..

Align seg 1/1 to: US-60-360-039-3962  from: 1 to: 639

473  CGCGCCGGGTGGCCGGGAAGTCTGTGTTCATTACGGGAGCTCACAGTC 424
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
319  ProAlaArgArgSerGlyLys.....LeuAspThrGluAsnHisGln.G 333
423  CTCCTCTTCGTCGCCAGTCGT..... 404
|||||  |||  |||  |||
333  luProAlaProProProArgPheAlaValProProIleAlaAspAla 349
403  .....CCTCGTCGTCTTCTTCAGCTCCCTCC 378
350  GlyLysPheAlaHisSerAspProProArgHisThrProSerAlaProG1 366

```



```
186 uArgLeuArgSerLeuLeuAlaAlaAspProAlaProLeuProProp 203
    ||| ||| |||
408 GTGCTCCGCGTGT .....CTTCTTCAGCTCCCTCGCGAGCGGTGCGG 365
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
203 roGlnProGlnGlnHisAlaLeuLeuHisGlyAlaProAlaAlaPro 219
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
364 GCG .....CGACAGACCGCGGTGGTGGAGCCGCGGATTCGAA 324
    ||| ||| |||
220 AlaGlyLeuThrLeuProPro ..... 227
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
323 CCCTGGCCATCGCTCCCGGACCCAGCGGGGCGCCACCGGTACCGA 274
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
228 .....ProProLeuProAspLysArgArgHisGlnHisProProProC 242
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
273 CGATCCGGCAACAG .....CGCGCGCATCACCGAGGGGGCAACCG 233
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
242 ysGlnGlnGlnGlnGlnGluProHisProAlaProGlnSerProLys 258
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
232 GCTTTTCATCAGCTGATGACATCAGCGCGTGGGTGTCACGACCCACT 183
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
259 AlaProThrAlaGluGluThrAlaAlaAla ..... 270
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
182 GCGCCAGTAGGACTCCGCGGAGCAGCGCGCGCGCGCGCTGGGGCC 133
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
271 .....AlaAlaGlnAlaAlaAlaAlaAlaAlaAla 280
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
132 TGATCCACGACGCGGATGTCGACAGCGGACTGTGCGGAGCAGGC 83
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
281 LysGluArgLysGluGluGlnArgArgLysGlnArgAspGluGluGlyLe 297
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
82 CCATCTCGCGGTCTCTGCTGCGGTGGTGGCGCGCGCGGTGCGCGGCC 33
    ||| ||| :|||:|||||:|||||:|||||:|||||:|||||:
297 uHisLeuLeuThrLeuLeuLeuGlnCysAlaGluAlaValAsnAlaAspA 314
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
32 ACCTGGCTGAACAGGAGTCACTGCTGCAG 1
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
314 snLeuAspAlaHisGlnThrLeuLeuGlu 324
```

seq\_name: /cgn2\_6/ptodata/2/paa/US10\_NEW\_COMB.pep:US-10-155-881-10047

seq documentation\_block:

; Sequence 10047, Application US/10155881

; GENERAL INFORMATION:

; APPLICANT: Dotson, Stanton B.

; APPLICANT: Kovalic, David K.

; APPLICANT: Liu, Jingdong

; APPLICANT: Lutfiyya, Linda L.

; APPLICANT: McIninch, James

; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH

; FILE REFERENCE: 38-21(15300)J

; CURRENT APPLICATION NUMBER: US/10/155,881

; CURRENT FILING DATE: 2002-05-22

; NUMBER OF SEQ ID NOS: 37595

; SEQ ID NO 10047

; LENGTH: 332

; TYPE: PRT

; ORGANISM: Zea mays

; US-10-155-881-10047

alignment\_scores:

Quality: 120.50 Length: 233

Ratio: 1.242 Gaps: 11

Percent similarity: 41.631 Percent identity: 25.751

alignment\_block:

US-09-462-480-2/rev x US-10-155-881-10047 ..

Align seg 1/1 to: US-10-155-881-10047 from: 1 to: 332

479 AGTCTCCGCGCGGTGGCGGG .....AAGTC 451

\* \* \*

```
||| |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
34 SerAlaProSerArgThrArgGlyArgArgGlyAlaAlaThrArgAl 50
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
450 TGTTCATATTACGGGAGCTCACAGTCCTCTTCCTCCAGTCGTCTCT 401
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
50 aSerThrThrThrProAlaProProAlaArgThrSerAlaArgAlaGly 67
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
400 GTCCTCTCTCTCA .....CGTCTCTGCGGAGCGGTGCCGCGC 363
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
67 laSerAlaAlaThrAlaAlaThrTrpArgThrProAlaSerGlyAlaGly 83
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
362 GCGACACACCGCGGTGGTGGAGCCCGGATTGCGAACCTGCGCCCAT 313
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
84 SerThrArgArgAlaThrAlaProSerProAlaArgThrAlaArgProAl 100
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
312 CGCTCCCGGACCCAGCGGAGCGGCCCGCTCACCGAC ..... 273
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
100 aAlaGlySerAlaSerSerArgThrProArgThrSerCysAlaCysS 117
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
272 .....GATCGGCAACAGCGCGCGCGCATC 249
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
117 erArgProArgSerSerSerProAlaProGlyAla.ProProAlaLe 133
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
248 ACCGAGGGGCAACCGGCTTTTCGATCAGTCGACATCAGCGCGGTGG 199
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
133 uProArg...SerProSerProThrThrAlaProArgSerGlyAlaArgA 149
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
198 GGTCAACGACCCACTGCGGAGGTAGGACTCCGCGCGCAGCAGCGCCG 149
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
149 rgSer.ArgAlaThr.SerProArgAlaSerCysAlaArgArgArgProA 165
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
148 CGCCC ..... 144
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
165 laProSerCysArgArgProSerArgProArgProArgSerProArgHisCys 181
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
143 .....GCGCTGGG .....CCTGATCCACAGCCAG 118
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
182 ArgProThrSerAlaValGlyAlaGlyArgAlaProGlyProProSerTh 198
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
117 CGGATGTTGTCAGCAGCGGACTGGTCCGCGAGCAGCGCCCATCTCGCGGCT 68
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
198 r .....ThrSerSerProArgSerAlaSerS 207
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
67 CCTCGTCGGTGGTGTG ..... 51
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
207 erAlaSerAlaGlyProThrArgProArgAlaGlyGlyAlaAlaThr 223
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
51 ..... 51
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
224 LeuHisProArgSerArgThrAspArgLeuArgGlyAlaArgSerThrAl 240
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
50 .....CGCGCGCGGTGCGCGCCACCTGG 27
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
240 aCysProProProArgGlyProProHisProAlaThrTrp 254
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

seq\_name: /cgn2\_6/ptodata/2/paa/US09\_NEW\_COMB.pep:US-09-573-655B-274

seq documentation\_block:

; Sequence 274, Application US/09573655B

; GENERAL INFORMATION:

; APPLICANT: SOLOVYEV, Victor and TROUKHAN, Maxim

; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptid

; TITLE OF INVENTION: Thereby

; FILE REFERENCE: 2750-0876P

; CURRENT APPLICATION NUMBER: US/09/573,655B

; CURRENT FILING DATE: 2000-05-18

; NUMBER OF SEQ ID NOS: 3281

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 274

; LENGTH: 312

; TYPE: PRT

; ORGANISM: Arabidopsis thaliana

; US-09-573-655B-274







OM of: US-09-462-480-2 to: PIR\_71.\* out\_format : pfs  
Date: Jul 22, 2002 1:26 AM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:  
-MODEL=frame+n2p.model -DEV=xlh  
-Q/Cgn2.1/USPTO.spool/US09462480/runat\_18072002\_164418\_19445/app\_query.fasta\_1.2850  
-DB=pir\_71 -QSWT=fastan -SUFFIX=rpr -CAPOP=12.000 -GAPEXT=4.000  
-MINWATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000 -GAPOP=4.500  
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -DELOP=6.000  
-DELEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0  
-ALIGN=15 -MODE=LOCAL -OUTFWT=pfs -NORM=ext -HEAPSIZE=500  
-MINLEN=0 -MAXLEN=200000000 -USER=US09462480\_@CGN1\_1\_73  
-NCPU=6 -ICPU=3 -LONGLOG -DEV\_TIMECUT=120 -WARN\_TIMECUT=30  
-NO\_XLPXY -WAIT -THREADS=1

Search information block:  
Query: US-09-462-480-2  
Query length: 524  
Database: PIR\_71.\*  
Database sequences: 283138  
Database length: 96089334  
Search time (sec): 87.620000

score_list:	Strd Orig	zScore	EScore	Len	Documentation
pir2:G70802	+ 713.00	936.10	1.3e-44	368	! probable PPE protein - Mycobact
pir2:T10033	+ 198.00	261.70	5.8e-07	302	! hypothetical protein MLCB628.14
pir2:S52796	- 140.00	183.48	0.0098	403	! prpL2 protein - human (fragment
pir2:S20590	- 133.00	168.34	0.0302	913	! exo-alpha-sialidase (EC 3.2.1.1
pir2:A46302	- 131.00	167.58	0.0430	707	! PTB-associated splicing factor,
pir2:T31611	- 131.00	161.70	0.0408	1585	! hypothetical protein Y50E8A.g
pir2:T22602	- 129.50	171.76	0.0595	304	! hypothetical protein F54B11.2
pir1:EM55	+ 127.00	160.90	0.0832	860	! elastin precursor - mouse
pir2:JC4163	+ 126.50	160.55	0.0908	825	! DNA-binding protein 5B5 - rat
pir2:D96711	- 125.50	160.35	0.1067	708	! hypothetical protein F24J5.8 [1
pir2:S20590	+ 125.50	158.50	0.1067	913	! exo-alpha-sialidase (EC 3.2.1.1
pir2:BA40505	+ 125.50	152.94	0.1014	1958	! hypothetical protein - suid he
pir2:T35474	- 125.00	162.90	0.1215	456	! 50kD proline rich protein - Str
pir2:A70507	- 124.50	157.44	0.1265	882	! probable-reductase (EC 1.3.99. -
pir2:T35192	- 124.00	158.02	0.1392	744	! probable ABC transporter - Stre
pir2:G70555	- 123.00	165.36	0.1781	227	! hypothetical protein Rv1158c -
pir2:S13367	- 123.00	158.20	0.1669	606	! Om(1D) protein - fruit fly (Dro
pir2:S06733	- 122.50	157.38	0.1813	620	! hydroxyproline-rich glycoprotein
pir2:T13078	- 122.50	155.78	0.1787	772	! KIAA0992 protein - human (fragm
pir2:T50568	- 121.50	150.49	0.2040	1334	! probable multi-domain regulato
pir2:S51939	- 121.00	157.93	0.2386	439	! chitinase (EC 3.2.1.14) precurs
pir2:B39369	- 120.00	155.07	0.2786	543	! homeotic protein BarH1 - fruit
pir2:A61183	- 119.50	158.45	0.3143	312	! hypothetical protein (sdsB regi
pir1:BA5344	+ 119.50	145.96	0.2807	1733	! probable nuclear antigen - sui
pir1:WMB338	+ 119.00	156.79	0.3388	358	! infected cell protein ICP34.5 -
pir2:H75457	- 119.00	156.07	0.3366	395	! hypothetical protein - Deinoco
pir2:T33663	- 119.00	155.56	0.3351	424	! hypothetical protein W04B5.3 -
pir2:TQ1696	- 119.00	155.52	0.3350	426	! pistil extensin-like protein pr
pir2:S49915	- 119.00	148.05	0.3131	1188	! extensin-like protein - maize
pir1:CGB025	+ 118.50	155.29	0.3658	402	! collagen alpha 2(I) chain - bov
pir2:T43556	- 118.50	152.69	0.3573	574	! Wiskott-Aldrich syndrome protei
pir2:T38819	- 118.50	152.69	0.3573	574	! Wiskott-Aldrich syndrome protei
pir2:A56154	- 118.50	151.42	0.3532	684	! Abi substrate ena (enabled) - H
pir2:T75311	+ 118.00	152.32	0.3897	552	! ABC transporter, ATP-binding pr
pir2:T04859	- 118.00	149.27	0.3791	839	! extensin homolog F28A21.80 - Ar
pir2:T46289	- 118.00	149.08	0.3784	862	! hypothetical protein DKF2p434A1
pir2:C86441	- 118.00	146.66	0.3702	1201	! unknown protein [imported] - A
pir2:A41724	- 118.00	146.59	0.3700	1213	! limb deformity (ld) protein - M
pir2:E70766	- 117.50	149.72	0.4165	721	! hypothetical protein Rv2082 - M
pir1:A45344	- 117.00	144.00	0.4327	1446	! immediate-early protein - suid

pir1:WMBEH6 - 117.00 138.29 0.4110 3164 ! UL36 protein - human herpes  
pir2:T20807 + 116.50 153.88 0.5179 340 ! hypothetical protein F13A7.1  
pir2:B70694 - 116.50 146.79 0.4857 900 ! probable infB - Mycobacteriu  
pir2:S54986 - 116.50 146.17 0.4830 980 ! regulatory protein - Emence  
pir2:JQ431 - 116.00 153.06 0.5625 348 ! hypothetical 35.5K protein -  
seq\_name: pir2:G70802  
seq\_documentation\_block:  
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)  
C.Species: Mycobacterium tuberculosis  
C.Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
C.Accession: G70802  
R.Cole, S.T.; Davies, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon,  
Connor, R.; Blosch, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A.Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A.Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno  
A.Reference number: A70500; MUID:98295987  
A.Accession: G70802  
A.Status: preliminary; nucleic acid sequence not shown; translation not shown  
A.Molecule type: DNA  
A.Residues: 1-368 <COL>  
A.Cross-references: GB:AL022120; GB:AL123456; NID:g3261558; PIDN:CAAL17965.1; PID:e126  
A.Experimental source: strain H37RV  
C.Genetics:  
A.Gene: PPE

alignment_scores:	Quality:	713.00	Length:	143
Percent Similarity:	Ratio:	5.057	Gaps:	0
US-09-462-480-2 x G70802	Percent Identity:	98.601	Percent Identity:	96.503
Align seg 1/1 to: G70802 from: 1 to: 368	alignment_block:			
1 CTCACAGGTGACGTCGTTGTTTCAGCCAGGTGGGGGGGACCGCGCGCGG 50				
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGly 242				
51 CAACCCAGCGCGAGGAGGAGCCGCGAGATGGCGCTCTCGGCACCAAGTC 100				
242 YASNProAlaAspGluGluAlaGlnMetGlyLeuGlyThrSerP 259				
101 CGCTGTGCAACCATCCGCTGGCTGGTGGATGATCAGGCCCCCAGCGCGCGCG 150				
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275				
151 GGCTGTGTCGCGCGGAGTCGCTACCTGGCGGAGGTGGGTGCTGTGACCCG 200				
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlySerLeuThrAr 292				
201 CACGCGCTCATGTCTCAGCTGATCGAAAGCCGGTTGCCCTCGGTGA 250				
292 gThrProLeuMetSerGlnLeulleGluLysProValAlaProSerValM 309				
251 TGCCGCGCGCTGTTCGGGATCGTTCGGTACGGGTGGCGCGCTCCGGTG 300				
309 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325				
301 GGTCCGGGACCGATGGCCAGGTTTGCATTCGCGGGGTCCACCAACCC 350				
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342				
351 GGCTGTGTCGCGCGCACCGCTCGCGGACGAGCGTGAAGAAAGACGACG 400				
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspG 359				
401 AGGACGACTGGGACGAGAGGACGACTGG 429				

|||||  
359 luAspAspTrpAspGluGluAspAspTrp 368

seq\_name: pir2:T10033

seq\_documentation\_block:

hypothetical protein MLCB628.14c - Mycobacterium leprae

C:Species: Mycobacterium leprae

C>Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 21-Jul-2000

C:Accession: T10033

R:Biglmeier, K.; Honore, N.; Woods, S.A.; Caudron, B.; Cole, S.T.

Mol. Microbiol. 7, 197-206, 1993

A:Title: Use of an ordered cosmid library to deduce the genomic organization of Mycobacterium leprae

A:Reference number: Z16917; MUID:93188700

A:Accession: T10033

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-302 <EIG>

A:Cross-references: EMBL:Y14967; NID:g2370268; PIDN:CAA75201.1; PID:g2370281

C:Genetics:

A:Note: MLCB628.14c

alignment\_scores:  
Quality: 198.00 Length: 126  
Ratio: 2.329 Gaps: 4  
Percent Similarity: 67.460 Percent Identity: 41.270

alignment\_block:

US-09-462-480-2 x T10033 ..

Align seg 1/1 to: T10033 from: 1 to: 302

1 CTGCAGCAGGTGACGCTGTTTTCAGCAGGTGGCGGCACCGCGCGG 50  
:|||||  
190 MetGlnValValysSerLeuPheThrSerIleAspSerThrGlyValty 206

51 CAACCCAGCCGAC.....GAGGAAGCCGCGCAGATGGCGCTGC 88  
::: |||||  
206 rThrSerAlaGlnArgGlyAspThrGluSerAlaHisArgIleGlyLeuP 223

89 TCGGCACAGTCGCTGCAACCATCCGTCGTCGTCGTCGTCGTCGTC 138  
:|||||  
223 heGlyAlaSerThrLeuSerSerHisProLeuValGlyIleThrGlyThr 239

139 AGCGCGGCGCGCGCGCTGTCGCGCGGAGTCGCTACCTGCGCAGTGG 188  
:|||||  
240 ThrThrAspThrArgLeuLeuCysAlaGluSerLeuProSerAlaSerGI 256

189 GTGCTTGACCCGCGCGCGCTGATGCT.....CAGCTGATCGAAAGC 232  
:|||||  
256 ySerLeuAlaTrpThrProLeuMetThrGlnPheGlnLeuIleAspLys 273

233 CGGTGCGCCCTCG.....GTGATGCGCGCGGTGTGTC 267  
:|||||  
273 erIleAlaProGluProArgGlnArgValMetLeuProTrpAlaAla 289

268 GGATGCGCGGTGACGGGTGCGCGCTGCGCTGCGTCCGGGAGCGATGG 317  
:|||||  
290 GlySerPro.....G 293

318 CCAGGGTTCGAATCGCGCGGTCCAC 345  
:|||||  
293 yHisAsnAlaGlnAspGlyGlyThrThr 302

seq\_name: pir2:S52796

seq\_documentation\_block:

priL2 protein - human (fragment)

C:Species: Homo sapiens (man)

C>Date: 19-May-1995 #sequence\_revision 10-Nov-1995 #text\_change 10-Sep-1997

C:Accession: S52796

R:Ruhmann, A.; Kreidewiss, S.; Nordheim, A.

submitted to the EMBL Data Library, March 1995

A:Reference number: S52796

A:Accession: S52796

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-403 <RUH>

A:Cross-references: EMBL:X86019; NID:g762950; PID:g762951

alignment\_scores:

Quality: 140.00 Length: 163

Ratio: 2.000 Gaps: 9

Percent Similarity: 42.945 Percent Identity: 32.515

alignment\_block:

US-09-462-480-2/rev x S52796 ..

Align seg 1/1 to: S52796 from: 1 to: 403

472 CGGCCCGGGTGGCGGGAAGTCTTGTTCATTACGGGAGCTCACCAGTCG 423  
|||||  
224 ArgProGlyProPro..... 228

422 TCCTCTTCGTCCTCCAGTCGT.....CCTC 400  
|||||  
229 .ProLeuProProSerSerSerGlyAsnAspGluThrProArgLeuProG 245

399 GTCGTCTTCTACGCTCCCTGCGGAGCGGTGCGCGGCGGACGAGCCG 350  
:|||||  
245 lnArgAsnLeu....SerLeuSerSerSerThrProLeuProSerPro 260

349 GGCTGGTGGAGCCCGCGGATTGCGAACCTTGCGCCATCGCTCCCGGACC 300  
|||  
261 Gly.....ArgSerGlyPro...LeuProProPr 269

299 ACCGAGCGGGCCACCGCTACCGCAGCATCCGCGGACGACGCGCGGCG 250  
:|||||  
269 oValProSerGluArgProProProValArgAspProGlyArgS 286

249 CA.....CCGAGGGGCAACCGCTTTTCGATCAGCTGAGACATCA 209  
|||  
286 erGlyProLeuProProProProValSerArgAsnGlySerThrSer 302

208 CGCGCGTGGGGTCAACGACCCACCTCGCCCA.....GGTAGCGAC 168  
|||  
303 ArgAlaLeuProAlaThr..ProGlnLeuProSerArgSerGlyValAsp 318

167 TCCGCGCGCAGCAGCGCGCGCGCTGGGGCTGATCCACCAGCCAG 118  
|||  
319 SerProArgSerGlyProArgProProLeuProProAspArgPro.... 333

117 CGGATGGTTCGACAGCGGACTGGTGCCGAGCAGGCCCATCTGCGCGGCTT 68  
333 .....

67 CCTCGTCGGTGGTTCGCCGCGCGGTGCGCGCCACC 30  
|||||  
334 .....SerAlaGlyAlaProProProProProSer 344

seq\_name: pir2:S20590

seq\_documentation\_block:

exo-alpha-sialidase (EC 3.2.1.18) - Actinomyces viscosus

C:Species: Actinomyces viscosus

C>Date: 22-Nov-1993 #sequence\_revision 01-Dec-1995 #text\_change 22-Oct-1999

C:Accession: S20590

R:Henningsen, M.; Roggentin, P.; Schauer, R.

Biol. Chem. Hoppe-Seyler 372, 1065-1072, 1991

A:Title: Cloning, sequencing and expression of the sialidase gene from Actinomyces vi

A:Reference number: S20590; MUID:92162190

A:Accession: S20590

A:Status: preliminary

A:Molecule type: DNA

A;Residues: 1-913 <HEI>  
A;Cross-references: EMBL:X62276; NID:g39254; PIDN:CAA44166.1; PID:g39255  
C;Keywords: glycosidase; hydrolase

alignment\_scores:  
Quality: 133.00 Length: 188  
Ratio: 1.529 Gaps: 9  
Percent Similarity: 46.277 Percent Identity: 29.787

alignment\_block:  
US-09-462-480-2/rev x S20590 ..

Align seg 1/1 to: S20590 from: 1 to: 913

```
472 CGGCGCGGTGGCGCGGAAGTCTGTGTCATTACGGGAGTCCACCAGTCG 423
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
686 ArgProArgArgProArgArgAlaLeuSerProArgArgHisArgHisH1 702
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
422 TCCTCTTCGTCCCGTCGTCCTCGTCGTCCTTCTTCACGCTCTCGCGGAG 373
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
702 sProArgProSerArgAlaLeuArg.....ProSerArgA 715
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
372 CGGTGCGCGGCG.....CGACCAGACCGGGGTGTGTGGAGC 338
|| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
715 laGlyProGlyAlaGlyAlaHisAspArgSerGluHisGlyAlaHisThr 731
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
337 CGCGGATTGCGACCTGCGCCCATCGTCCCGGACCCACCGAGCGGCG 288
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
732 GlySerCysAlaGlnSerAlaProGluGlnThrAspGlyProThrAlaA1 748
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
287 CCACCGCTACCGACGATCCCGGCAACAGCCGCGGCATCACCGAGGGGCG 238
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
748 a...ProAlaProGluThrSerSerAlaProAlaAlaGluProThrGlnA 764
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
237 AACCGGCTTTTCGA.....TCAGCTGAG 215
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
764 laProThrValAlaProSerValGluProThrGlnAlaProGlyAlaGln 780
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
214 ACATCAGCGGCG.....TGCGGTTCACAGCCACCTGCGCCAGGTAGC 171
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
781 ProSerSerAlaProLysProGlyAlaThrGlyArgAlaProSerValva 797
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
170 GACTCCGCGCGCAGCAGGC..... 152
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
797 laSnProLysAlaThrGlyAlaAlaThrGluProGlyThrProSerSers 814
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
151 .....CGCGCGCGCGTGGGCG 134
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
814 erAlaSerProAlaProSerArgAsnAlaAlaProThrProLysProGly 830
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
133 CTG.....ATCCACAGCCAGCGGATGTTCCAGCAGCGGACTG 96
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
831 MetGluProAspGluIleAspArgProSerAspGlyThrMetAlaGlnPr 847
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
95 GTGCGGACGAGCCCATCTGCGCGCTTCCTCGTGGCTGGGTGGCGGCC 46
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
847 oThrGlyAla...ProAlaArgValProArg.....ArgA 860
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
45 GCCGGTGGCGGCCA 32
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
860 rGArgArgArgPro 864
```

seq\_name: pir2:A46302

seq\_documentation\_block:

PTB-associated splicing factor, long form - human  
N;Alternate names: myoblast cell surface antigen 24.1D5; polypyrimidine tract-binding protein  
C;Species: Homo sapiens (man)  
C;Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 24-Sep-1999  
C;Accession: A46302; A43557; S29995  
R;Patton, J.G.; Porro, E.B.; Galceran, J.; Tempst, P.; Nadal-Ginard, B.  
Genes Dev. 7, 393-406, 1993

A;Title: Cloning and characterization of PSF, a novel pre-mRNA splicing factor.  
A;Reference number: A46302; MUID:93194059  
A;Accession: A46302  
A;Molecule type: mRNA  
A;Residues: 1-707 <PAT>  
A;Cross-references: EMBL:X70944; NID:g38457; PIDN:CAA50283.1; PID:g38458  
A;Note: sequence extracted from NCBI backbone (NCBIP:127206)  
R;Gower, H.J.; Moore, S.E.; Dickson, G.; Elsom, V.L.; Nayak, R.; Walsh, F.S.  
Development 105, 723-731, 1989  
A;Title: Cloning and characterization of a myoblast cell surface antigen defined by  
A;Reference number: A43557; MUID:90091812  
A;Accession: A43557  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 312-707 <GOW>  
A;Cross-references: GB:X16850; NID:g23711; PIDN:CAA34747.1; PID:g23712  
C;Genetics:  
A;Gene: GDB:SFPQ; PSF  
A;Cross-references: GDB:138275  
A;Map position: 4q-4q  
C;Superfamily: unassigned ribonucleoprotein repeat-containing proteins; ribonucleoprotein repeat  
C;Keywords: alternative splicing; pre-mRNA splicing; surface antigen  
F;298-359/Domain: ribonucleoprotein repeat homology <RRM1>  
F;372-438/Domain: ribonucleoprotein repeat homology <RRM2>

alignment\_scores:  
Quality: 131.00 Length: 139  
Ratio: 1.899 Gaps: 9  
Percent Similarity: 49.640 Percent Identity: 31.655

alignment\_block:

US-09-462-480-2/rev x A46302 ..

Align seg 1/1 to: A46302 from: 1 to: 707

```
367 CCGCGCGCAGCAGACCCGCGGTGTGTGGAGCCGCGGATTGCGAACCTGG 318
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
46 ProMetGlyProGlyProGlyGln.....SerGlyProly 57
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
317 CCCATCGTCCGCGACCCACCG.....GAGCGCGCCACCGGTAC 277
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
57 sProProileProProProProProHisGlnGlnGlnGlnProProp 74
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
276 CGACGATCCGCGCACAGCCGCGCATCA.....CCGAGGGGGCAACCG 233
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
74 ro.....GlnGlnProProGlnGlnProProHisGlnPro 87
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
232 GCTTTTCATCAGTCAGACATCAGCGGCGTGGCGGTCAACGACCCACT 183
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
88 ProProHisProGlnPro.HisGlnGln.....GlnGln.ProPro 100
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
182 GCGCCAGGTAGCTCCGCGCGC.....AGCAGGCC 151
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
101 ProProProGlnAspSerSerLysProValValAlaGlnGlyProGlyPr 117
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
150 CGCGCGCGCGTGGGGCTGATCCACCAGCCAGCGGA..... 114
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
117 oAlaProGlyValGlySerAlaProProAlaSerSerSerAlaProProA 134
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
113 .....TGGTTCGACAGCGGACTGTGTG 93
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
134 laThrProProThrSerGlyAlaProProGlySerGlyProGlyProThr 150
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
92 CCGAGCAGGCCCATCTGCGCGGCTTCCTCGTGGGTGGTTCGCCGCCGCC 43
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
151 ProThrProProProAlaValThrSerAlaProProGlyAlaProProPr 167
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
42 GGTGCCGCCGCCACC 30
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
167 oThrProProSer 171
```

seq\_name: pir2:T31611



seq\_name: pir1:EAMS

seq\_documentation\_block:  
 elastin precursor - mouse  
 N:Alternate names: tropoelastin  
 C:Species: Mus musculus (house mouse)  
 C>Date: 18-Aug-1995 #sequence\_revision 16-Aug-1996 #text\_change 22-Jun-1999  
 C:Accession: A55721  
 R:Wyder, K.S.; Sechler, J.L.; Boyd, C.D.; Passmore, H.C.  
 Genomics 23, 125-131, 1994  
 A:Title: Use of an intron length polymorphism to localize the tropoelastin gene to mouse chromosome 12  
 A:Reference number: A55721; MUID:95130069  
 A:Accession: A55721  
 A:Molecule type: mRNA  
 A:Residues: 1-860 <WVD>  
 A:Cross-references: GB:U08210; NID:9473273; PIDN:AMA80155.1; PID:9473274  
 C:Genetics:  
 A:Map position: 5  
 C:Superfamily: elastin  
 C:Keywords: alternative splicing; extracellular matrix; glycoprotein; hydroxyllysine  
 F:1-27/Domain: signal sequence #status predicted <SIG>  
 F:28-860/Product: elastin #status predicted <WAT>  
 F:850-855/Disulfide bonds: #status predicted

alignment\_scores:  
 Quality: 127.00 Length: 141  
 Ratio: 1.649 Gaps: 8  
 Percent Similarity: 54.610 Percent Identity: 34.752

alignment\_block:

US-09-462-480-2 x EAMS ..

Align seg 1/1 to: EAMS from: 1 to: 860

31 GTGGGGGCGACCGG.....GGCGCAACCC 56  
 ::::: |||  
 53 IleglyGlyLeuGlyGlyGlyAlaLeuGlyProGlyGlyLysPr 69  
 57 AGCGGACGAGGAGCGCGCAGAGTGGCTCTCGGCACCACTCGCTGT 106  
 69 oProLysProGlyAlaGlyLeuLeuGlyThrPheGlyAlaGlyProGlyG 86  
 107 CGAACCATCCGCTGGCTGGATCAGCCCGCGCGCGCGCGCGCTG 156  
 ::::: |||  
 86 ly.....LeuGlyGlyAlaGlyProGlyAlaGlyLeuGlyAla 98  
 157 CTGCGCGGAGTCCGTACCTACGCGCGAGTGGTGGTGTGACCG..... 200  
 ::::: |||  
 99 PheProAlaGlyThrPheProGlyAlaGlyAlaLeuVal..ProGlyGly 114  
 201 .....CACGCGGTGTCTCAGCTGATCGAAAGCGGTGCCC 241  
 ::::: |||  
 115 AlaAlaGlyAlaAlaAlaTyrLysAlaAlaLysAlaGlyAlaG 131  
 242 CCTCGGTGATCCGCGCTGTTCGCGA.....TCGTCGGTG 279  
 ::::: |||  
 131 yLeuGlyGlyValGlyGlyValProGlyGlyValGlyGlyVal 147  
 280 ACGGTGGCGCGCTCCGGTGGGT...CCGGAGCATGGGCCAGGTC 326  
 ::::: |||  
 148 ProGlyGlyValGlyValGlyGlyValProGlyGlyValGlyGly 164  
 327 GCATCCGCGCGC.....TCCACCA 346  
 ::::: |||  
 164 yValProGlyGlyValGlyGlyIleGlyGlyIleGlyGlyVal 181  
 347 GCCCGGCTCGCGCGC 366  
 ::::: |||  
 181 erThrGlyAlaValPro 187

seq\_name: pir2:JC4163

seq\_documentation\_block:  
 DNA-binding protein 5E5 - rat  
 C:Species: Rattus norvegicus (Norway rat)  
 C>Date: 27-Aug-1995 #sequence\_revision 19-Oct-1995 #text\_change 20-Jun-2000  
 C:Accession: JC4163; PC4040  
 R:Suzuki, E.; Kojima, N.; Yoshimura, K.; Obata, K.; Akagawa, K.  
 J. Biochem. 118, 122-128, 1995  
 A:Title: Cloning and sequence analysis of cDNA for a possible DNA-binding protein 5E1  
 A:Reference number: JC4163; MUID:96015159  
 A:Accession: JC4163  
 A:Molecule type: mRNA  
 A:Residues: 1-825 <SUZ>  
 A:Cross-references: DDBJ:D37934; NID:g531260; PIDN:BAA07153.1; PID:g531261  
 A:Experimental source: brain  
 A:Accession: PC4040  
 A:Molecule type: protein  
 A:Residues: 230-455 <SUZ>  
 C:Comment: This protein has an abundance of arginine, a glycine-rich region and a proline-rich region  
 C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homol  
 C:Keywords: nerve; phosphoprotein  
 F:436-443/Region: nuclear location signal  
 F:722-731/Region: proline cluster  
 F:62,258,345,360,404,413,570,613,635,752,820/Binding site: phosphate (Ser) (covalent)

alignment\_scores:  
 Quality: 126.50 Length: 183  
 Ratio: 1.265 Gaps: 8  
 Percent Similarity: 54.645 Percent Identity: 32.240

alignment\_block:

US-09-462-480-2 x JC4163 ..

Align seg 1/1 to: JC4163 from: 1 to: 825

24 CAGCCAGGTGG.....CGGCACCGCGCGCGCAACCCAGCAGAGG 67  
 478 GlnAlaGlyGlyGlyProArgGlyArgAlaArgGlyProArgGlnGlnAl 494  
 68 AAGCGCGCAGATGGCTGCTCGGCACCACTCGCTGTCGAACATCCG 117  
 494 aArgArgHisGlyProGlnArgArgGly.....Pro.Pro 507  
 118 CTGCTTGTGTGATCAGCCCGCGCGCGCGCTCTG..... 159  
 ::::: |||  
 508 GlnAlaGlyGluGlyProGlyAspAlaThrLeuValLeuGlyLeuG 524  
 160 .....CGCGCGAGTCGTACTGCGC 181  
 524 yThrThrSerGlyGluGlnArgAlaAspGlnSerGlnThrLeuProAla 541  
 182 CAGTGGGTCTTACCCG.CACGCCCTGATGTCAGTCAGTCGAGAAA 230  
 ::::: |||  
 541 euAlaGlyAlaProThrAlaHisAlaHisAlaValProGlyProGlyPro 557  
 231 GCCGTTTGCCTCCCTCGGTGTCGCGC.....GGCTGTCGCGATCGT 274  
 ::::: |||  
 558 AlaAlaThrLeuGlyGlyArgGlyArgGlySerTrpArg..... 572  
 275 CGGTGACGGGTGGCGCTCCGGTGGGTCCGGAGGAGGATGGCCAGG 324  
 ::::: |||  
 573 .....GlyGlyArgGlyGlyGlyAlaGlyAlaSerGlyGlyAla 587  
 325 TCGCAATCCCG.....CGGTCACCAACCGCGCGGTCTGTGCGGCC 365  
 ::::: |||  
 587 rgGlyGlyArgGlyArgGlyGlyArgGlyArgGlySerGlyLeuSer 603  
 366 GGCACCTCCG...GCAGGACCTGAAGAGACGACGACGACCTGG 412  
 ::::: |||  
 604 GlyThrArgGluAspAlaGlySerProSerAlaArgGlyGlyGlnAr 620  
 413 ACGAAGAGGACGACTGGTGGTGGTCCCGTAATGACAACAGACTTCCCGGCC 462

|||||.....  
620 gArgGlyHisGlyProProAlaAlaGlyAlaGlnValSerThrA 637  
463 ACCCGCGCGGAAGATTGCGCAACATTTTGGCGAGGAGGTAAAGAG 509  
637 rGlyArgAlaAlaArgGlyGlnArgThrGlyGluGluAlaGlnasp 652  
seq\_name: pir2:D96711

seq\_documentation\_block:  
hypothetical protein F24J5.8 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
R:Accession: D96711  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719  
A:Accession: D96711  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-708 <STO>  
A:Cross-references: GB:AE005173; NID:g5734709; PIDN:AAD49974.1; GSPDB:GN00141  
C:Genetics:  
A:Gene: F24J5.8  
A:Map position: 1

alignment\_scores:  
Quality: 125.50 Length: 175  
Ratio: 1.793 Gaps: 7  
Percent Similarity: 40.000 Percent Identity: 29.143

alignment\_block:  
US-09-462-480-2/rev x D96711 ..

Align seg 1/1 to: D96711 from: 1 to: 708

421 CCTCTGCTCCAGTCGTCCTCGTCGTCCTTTCACGCTCTCTGCGCAGC 372  
|||||.....  
38 ProLeuProProSerAlaPro.....ProProAsnArgAl 49  
371 GGTCGCGCGGACACCGCGGCTGGTGAGCCGCGCGGATTCGGAAC. 323  
| |||.....  
49 aProProProProProProValThrThrSerProProProValAlaAsnG 66  
322 .....CTGGCCCATCGCTCCCGACCCACCGGAGCGCGCCAC..... 284  
|||||.....  
66 lAlaProProProLeuProProProProProGluSerSerProPro 82  
283 CCCTCACCGAGATCCCGCAACACCGCGCGCA...TCACGAGGGGCA 237  
|||||.....  
83 ProGlnProValIleProSerProProProSerThrSerProProG 99  
236 ACCGGCTTTTCATCAGCTCAGACATCAGCGCGGTG..... 201  
|||||.....  
99 nProValIleProSerProProProProSerAlaSerProProAlaLeuV 116  
201 .....  
116 alProProLeuProSerSerProProProProAlaSerValProProPro 132  
201 .....  
133 ArgProSerProSerProProIleLeuValValArgSerProProProSerVa 149

200 .....CGGTCAACGACCCACCTGCGCCAGGTAGCAGCTCCGCGCGCAGCA 155  
:.....  
149 LArgProIleGlnSerProProProProSerAspArgProThrGlns 166  
154 GGCCCGCGCGCGCTGGGCTGATCCACGACCGCGGATGTTGCGAC 105  
||| ||| .....  
166 erProProProProSerProProSerProProSerGluArgProThrGln 182  
104 AGCGGACTGGTGGCGAGCAGCGCATCTGCGGCTTCTCGTCGGCTGG 55  
||| |||.....  
183 Ser.....ProProSerProProSerGluArgProThrGlnSerPro... 196  
54 GTTCCGCGCGCGCTGGCGCCAC 30  
|||||.....  
197 ....ProProProSerProProSer 203  
seq\_name: pir2:S20590  
seq\_documentation\_block:  
exo-alpha-sialidase (EC 3.2.1.18) - Actinomyces viscosus  
C:Species: Actinomyces viscosus  
C:Date: 22-Nov-1993 #sequence\_revision 01-Dec-1995 #text\_change 22-Oct-1999  
C:Accession: S20590  
R:Henningsen, M.; Roggentin, P.; Schauer, R.  
Biol. Chem. Hoppe-Seyler 372, 1065-1072, 1991  
A:title: Cloning, sequencing and expression of the sialidase gene from Actinomyces vi  
A:Reference number: S20590; MUID:92162190  
A:Accession: S20590  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-913 <HEI>  
A:Cross-references: EMBL:X62276; NID:g39254; PIDN:CAA44166.1; PID:g39255  
C:Keywords: glycosidase; hydrolase

alignment\_scores:  
Quality: 125.50 Length: 204  
Ratio: 1.268 Gaps: 9  
Percent Similarity: 48.529 Percent Identity: 28.431

alignment\_block:  
US-09-462-480-2 x S20590 ..

Align seg 1/1 to: S20590 from: 1 to: 913

36 CGGACCGCGCGCAACCCAGCAGCAGGAGCGCGCAGATGGGCC 85  
|||||.....  
680 ArgHisArgArgArgSerArgProArgArgProArgAlaLeuSerPr 696  
86 TGCTGGCACCAG.....T 99  
| |||.....  
696 oArgArgHisArgHisHisProProArgProProSerArgAlaLeuArgProS 713  
100 CCCTGTGCAACCATCCGCTGGCTGGTGGATCAGG..... 134  
|| .....  
713 erArgAlaGlyProGlyAlaGlyAlaHisAspArgSerGluHisGlyAla 729  
135 .....CCCGCGCGCGCGCGCGCGCTGC 157  
|||||.....  
730 HisThrGlySerCysAlaGlnSerAlaProGluGlnThrAspGlyProTh 746  
158 TGC CGCG.....GAGTCGTACCTGGCGCAGGTGGTGGTTCGAC 198  
|||||.....  
746 rAlaAlaProAlaProGluThrSerSerAlaProAlaAlaGluProThr 762  
199 CGCAGCGCGCTGATGCTCAGCTGATCGAAAGCGGTTGCCCTCGGT 248  
:.....  
763 GlnAlaProThrValAlaProSerValGluProThrGlnAlaProGlyAl 779  
249 GATCGCGCGCGCTGTCGCCGATCGTCGTCACGGGTGGCGCC...GCTC 295  
:.....  
779 aGlnProSerSerAlaProLysProGlyAlaThrGlyArgAlaProServ 796







```

437 GGAGCTCACCAGTGTCTCTCTGTCCTCCAGTCGTCTCTCTGTCGTCTCTCTC 388
||||| : : : : : ||| : : : : : : : : : :
293 GlyAlaGlyGluIleProAlaAlaArgGlnGlnGluArgArgGluAlaAs 309
||||| : : : : : ||| : : : : : : : : : :
387 ACCTCTCTGCGGAGCGGTCCGGC.....GCGACGAGACCGGGC 347
: : : ||| ||| : : : ||| ||| ||| ||| |||
309 pGlyGluArgAlaProGlyThrGlyThrLeuSerAlaGlyArgProGly. 325
||||| : : : : : ||| : : : : : : : : : :
346 TGTGTGAGCGCGCGGATTGCGAACCCTGCGCCCATCGTCCCGGACCCACC 297
||||| : : : : : ||| : : : : : ||| : : : : :
326 .....ProGlu...ThrProGlyValLeuLeuProSerProAla 337
296 GGAGCGCGCCACCGGTACCGACGATCCGGCAACAGCCCGCGCATCAC 247
: : : ||| : : : ||| : : : : : |||
338 ThrGluGlyGluProLeuGluAspGlyArgGlyAlaGlyHisGlyAspG1 354
||||| : : : : : ||| : : : : : |||
246 CGAGGGGGCAACCGGCTTTTCGATCAGCTGAGACATCAGCGGGGTGCGGG 197
: : : ||| : : : : : ||| : : : : : |||
354 yAspGlyProAlaSerValAlaValProAsnArgThrProGlyProArg. 370
||||| : : : : : ||| : : : : : |||
196 TCAACGACCCACCTGCGCCA.....GGTAGCGACTCCGCGCGC 159
: : : ||| ||| ||| ||| ||| : : : : : |||
371 .....GlnAlaProAlaProValSerGlyHisGlyProGluAlaAlaPro 385
158 AGCAGGCGCGCGCGCGCTGGGGCTGTATCCACCGAGCGGATGGTT 109
||| ||| ||| ||| ||| ||| ||| ||| |||
386 SerProSerAlaProAlaProGlyProSerGluProAlaSerGly..... 400
108 CGACAGCGGACTGTGCGCGAGAGGCCCATCTGCGGGCTTCTCTCGTCGG 59
||||| : : : : : ||| : : : : : |||
401 .ProSerAlaProAlaProGlyProProAlaProAlaAlaGlyProSerA 417
58 CTGGGTTGCCCGCGCGGTGCGGCC 33
|| ||| ||| |||
417 laProAlaProGlyProSerAlaPro 425

```

**THIS PAGE BLANK (USPTO)**





```

107 aProSerProAlaProSerProAla.....ProSer 120
252 CATCACGAGGCGCAACCGCTTTTCAGTCAGCTGAGAGCATCAGCGCG 203
120 erProSerProAlaProPro..SerProSer..... 130
202 TGCAGGTCACAGCACCCACCTCGCCAGTAGAGCTCCGCGCCGACGAGG 153
131 .....ProProAlaProProSerProSerProSer... 141
152 CCGCGCGCGCGCTGGGCGCTGATCCA....CCAGCCAGCGGATGGTTCGA 106
142 ProAlaProProProProProSerProAlaProSer..... 154
105 CAGCGGACTGGTCGAGGAGCGCCATCTCGCGGCTTCCTCGTCGCGCTG 56
155 .....ProSerProProValProProSerProSer..... 164
55 GGTTCGCGCGCGCGTGGCGCCGCCACC 30
165 .....ProProValProProSer 170

```

seq\_name: SwissProt\_40:SFQ\_HUMAN

seq\_documentation\_block:

```

ID SFQ_HUMAN STANDARD; PRT; 707 AA.
AC P23246; P30808;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Splicing factor, proline and glutamine-rich (Polyproline tract-
DE binding protein-associated splicing factor) (PTB-associated splicing
DE factor) (PSF) (DNA-binding P52/P100 complex, 100 kDa subunit).
GN SFQ OR PSF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND ALTERNATIVE SPLICING.
RC TISSUE=Fetal brain;
RX MEDLINE=93194059; PubMed=8449401;
RA Patton J.G., Porro E.B., Galceran J., Tempst P., Nadal-Ginard B.;
RT "Cloning and characterization of PSF, a novel pre-mRNA splicing
RT factor."
RL Genes Dev. 7:393-406(1993).
RN [2]
RP SEQUENCE OF 312-707 FROM N.A.
RC TISSUE=Fetal skeletal muscle;
RX MEDLINE=90091812; PubMed=2480877;
RA Gower H.J., Moore S.E., Dickson G., Elsom V.L., Nayak R., Walsh F.S.;
RT "Cloning and characterization of a myoblast cell surface antigen
RT defined by 24 LD5 monoclonal antibody."
RL Development 105:723-731(1989).
RN [3]
RP SEQUENCE OF 48-68 AND 213-246.
RX MEDLINE=93176127; PubMed=8439294;
RA Zhang W.-W., Zhang L.-X., Busch R.K., Farres J., Busch H.;
RT "Purification and characterization of a DNA-binding heterodimer of 52
RT and 100 kDa from HeLa cells."
RL Biochem. J. 290:267-272(1993).
CC -1- FUNCTION: ESSENTIAL PRE-MRNA SPLICING FACTOR REQUIRED EARLY IN
CC SPICEOSOME FORMATION. BINDS TO THE MAMMALIAN POLYPYRIMIDINE
CC TRACTS. FORMS A COMPLEX WITH THE POLYPYRIMIDINE TRACT-BINDING
CC PROTEIN (PTB). SEEMS TO ALSO BIND DNA.
CC -1- SUBUNIT: HETEROTETRAMER OF TWO 52 KDA AND TWO 100 KDA SUBUNITS.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; A LONG FORM (SHOWN
CC HERE) AND A SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: CONTAINS 2 RNA RECOGNITION MOTIFS (RRM).
CC -1- CAUTION: WAS ORIGINALLY (REF.2) THOUGHT TO BE MYOBLAST CELL
CC SURFACE ANTIGEN 24.LD5 AND A POSSIBLE MEMBRANE-BOUND PROTEIN

```

ECTOKINASE.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----

```

DR EMBL; X70944; CAA50283.1; -
DR EMBL; X16850; CAA34747.1; -
DR PIR; A43557; A43557.
DR PIR; S29770; S29770.
DR HSP; P19339; 1SXL.
DR MIM; 605199; -
DR InterPro; IPR000504; RRM.
DR Pfam; PF00076; rrm; 2.
DR SMART; SM00360; RRM; 2.
DR PROSITE; PS0102; RRM; 2.
DR PROSITE; PS00030; RRM_RNP_1; 1.
KW Nuclear protein; RNA-binding; DNA-binding; mRNA splicing; Repeat;
KW Alternative splicing
FT DOMAIN 297 369 RNA-BINDING (RRM) 1.
FT DOMAIN 371 452 RNA-BINDING (RRM) 2.
FT REPEAT 9 27 3 X 3 AA REPEATS OF R-G-G.
FT REPEAT 19 11 1.
FT REPEAT 25 27 2.
FT DOMAIN 10 266 GLN/GLU/PRO-RICH.
FT DOMAIN 10 15 POLY-GLY.
FT DOMAIN 20 27 POLY-GLY.
FT DOMAIN 56 65 POLY-PRO.
FT DOMAIN 67 71 POLY-GLN.
FT DOMAIN 95 98 POLY-GLN.
FT DOMAIN 99 103 POLY-PRO.
FT DOMAIN 184 188 POLY-PRO.
FT DOMAIN 571 574 POLY-ARG.
FT DOMAIN 613 616 POLY-GLY.
FT DOMAIN 635 641 POLY-GLY.
FT VARSPIC 663 707
FT CONFLICT 243 243 G -> R (IN REF. 3).
FT SEQUENCE 707 AA; 76149 MW; 6D8D5EA95E235847 CRC64;
SQ

```

alignment\_scores:

Quality: 131.00 Length: 139  
 Ratio: 1.899 Gaps: 9  
 Percent Similarity: 49.640 Percent Identity: 31.655

alignment\_block:

US-09-462-480-2/rev x SFQ\_HUMAN ..

Align seg 1/1 to: SFQ\_HUMAN from: 1 to: 707

```

367 CCGCGCGGACGACACCGCGGCTGGTGAGCGCGCGGATTCGAAACCTGG 318
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
46 ProMetGlyProGlyProGlyGln.....SerGlyProLy 57
317 CCCATCTCTCCCGGACCCACCG.....GAGCGCGGCGCCGCTAC 277
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
57 sProProileProProProProHisGlnGlnGlnGlnGlnProProp 74
276 CGAGCATCCGCGACACGCGCGCGCATCA.....CCGAGGGGCAACCG 233
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
74 ro.....GlnGlnProProProGlnGlnGlnGlnGlnProHisGlnPro 87
232 GCTTTTCATCAGTCAGATCAGCGCGGTGCGGGTCAACGACCCACCT 183
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
88 ProProHisProGlnProHisGlnGln.....GlnGlnProPro 100
182 GCGCCAGTAGCAGTCCGCGCGC.....AGCAGGCC 151

```

```

101 ProProGlnAspSerSerLysProValValAlaGlnGlyProGlyPr 117
150 CGCGCCGCGCTGGCGGCTGATCCACGAGCAGCGGA..... 114
117 oAlaProGlyValGlySerAlaProProAlaSerSerAlaProProA 134
113 .....TGGTTCGACAGCGGACTGGTG 93
134 laThrProProThrSerGlyAlaProProGlySerGlyProGlyProThr 150
92 CCGAGCAGGCGCATCTCGCGGCTTCCTCGCTCGGCTGGGTGCGCGGCC 43
151 ProThrProProProAlaValThrSerAlaProProGlyAlaProProPr 167
42 GGTGCGCGCCACC 30
167 oThrProProSer 171

```

seq\_name: SwissProt\_40:ELS\_MOUSE

```

seq_documentation_block:
ID ELS_MOUSE STANDARD; PRT; 860 AA.
AC P54320.
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Elastin precursor (Tropoelastin).
GN ELN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C; TISSUE=Lung;
RX MEDLINE=95130069; PubMed=7829060;
RA Wyder K.S., Sechler J.L., Boyd C.D., Passmore H.C.;
RT "Use of an intron polymorphism to localize the tropoelastin gene to
RT mouse chromosome 5 in a region of linkage conservation with human
RT chromosome 7.";
RL Genomics 23:125-131(1994).
CC -!- FUNCTION: MAJOR STRUCTURAL PROTEIN OF TISSUES SUCH AS AORTA AND
CC NUCAL LIGAMENT, WHICH MUST EXPAND RAPIDLY AND RECOVER COMPLETELY.
CC -!- SUBUNIT: THE POLYMERIC ELASTIN CHAINS ARE CROSS-LINKED TOGETHER
CC INTO AN EXTENSIBLE 3D NETWORK.
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX OF ELASTIC FIBERS.
CC -!- PTM: THE CROSSLINKS ARE MADE OF DEAMINATED LYS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U08210; AAR80155.1; .
DR HSSP; P04002; LWFA.
DR MGD; MGI:95317; Eln.
DR InterPro; IPR003979; tropoelastin.
DR PRINTS; PR01500; TROPOELASTIN.
KW Structural protein; Repeat; Signal; Connective tissue.
FT SIGNAL 1 27 POTENTIAL.
FT CHAIN 28 860 ELASTIN.
SQ SEQUENCE 860 AA; 71955 MW; 0C0BESAEE1EDD7F1 CRC64;

```

alignment\_scores:  
 Quality: 127.00 Length: 141  
 Ratio: 1.649 Gaps: 8  
 Percent Similarity: 54.610 Percent Identity: 34.752

alignment\_block:  
 US-09-462-480-2 x ELS\_MOUSE ..

Align seg 1/1 to: ELS\_MOUSE from: 1 to: 860

```

31 GTGGGCGCGCACCGCG.....GGCGGCAACCC 56
   ::::: ||| ||| ||| ||| |||
53 ileGlyGlyLeuGlyGlyGlyGlyGlyAlaLeuGlyProGlyGlyLysPr 69
   | | | | | | | | | | | | | | | | | |
57 AGCGCAGCAGGAAGCGCGCAGATGCTCGGCACCGACGTCGCGTGT 106
   | | | | | | | | | | | | | | | | | |
69 oProLysProGlyAlaGlyLeuGlyThrPheGlyAlaGlyProGlyG 86
   | | | | | | | | | | | | | | | | | |
107 CGAACCATCCGCTGGCTGGTGGATCAGCCCGCGCGGCGCGGCGCTG 156
   :: | | | | | | | | | | | | | | | | |
86 ly.....LeuGlyGlyAlaGlyProGlyAlaGlyLeuGlyAla 98
   | | | | | | | | | | | | | | | | | |
157 CTGCGCGCGGAGTCGTACCTGGCGCAGGTGGTGGTGGTACCCG..... 200
   :: | | | | | | | | | | | | | | | | |
99 PheProAlaGlyThrPheProGlyAlaGlyAlaLeuVal..ProGlyGly 114
   | | | | | | | | | | | | | | | | | |
201 .....CACGCGCGCTGATCTCAGCTGATCGAAAAGCGGTGGCC 241
   | | | | | | | | | | | | | | | | | |
115 AlaAlaGlyAlaAlaAlaAlaTyrLysAlaAlaLysAlaGlyAlaGl 131
   | | | | | | | | | | | | | | | | | |
242 CCTCGGTGATCGCGCGGCTGTTCGCGGA.....TCGTGGGTG 279
   | | | | | | | | | | | | | | | | | |
131 yLeuGlyGlyValGlyGly..ValProGlyGlyValGlyValGlyVal 147
   | | | | | | | | | | | | | | | | | |
280 ACGGTGCGCGCGCTCGGTGGGT....CCGGCAGCGATGGCCAGGGTTC 326
   | | | | | | | | | | | | | | | | | |
148 ProGlyGlyValGlyValGlyValProGlyGlyValGlyValGlyGl 164
   | | | | | | | | | | | | | | | | | |
327 GCAATCCGCGCGC.....TCCACCA 346
   : | | | | | | | | | | | | | | | | |
164 yValProGlyGlyValGlyGlyIleGlyGlyLeuGlyLeuGlyValS 181
   | | | | | | | | | | | | | | | | | |
347 GCCCGGTGCTGGTCGCGCGCG 366
   | | | | | | | | | | | | | | | | | |
181 erThrGlyAlaValValpro 187

```

seq\_name: SwissProt\_40:5E5\_RAT

seq\_documentation\_block:

```

ID 5E5_RAT STANDARD; PRT; 825 AA.
AC Q63003.
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE 5E5 antigen.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar; TISSUE=Brain;
RX MEDLINE=96015159; PubMed=8537300;
RA Suzuki E., Kojima N., Yoshimura K., Uyemura K., Obata K., Akagawa K.;
RT "Cloning and sequence analysis of cDNA for a possible DNA-binding
RL protein 5E5 in the nervous system.";
RL J. Biochem. 118:122-128(1995).
CC -!- FUNCTION: MIGHT HAVE DNA-BINDING ABILITY.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN NEURONS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```



CC -----  
DR EMBL; D37934; BAA07153.1; ..  
KW DNA-binding; Nuclear protein; Antigen.  
SQ SEQUENCE 825 AA; 86831 MW; AF667FE2FD555BDF CRC64;

alignment\_scores:  
Quality: 126.50 Length: 183  
Ratio: 1.265 Gaps: 8  
Percent Similarity: 54.645 Percent Identity: 32.240

alignment\_block:

US-09-462-480-2 x 5E5\_RAT

Align seg 1/1 to: 5E5\_RAT from: 1 to: 825

```
24 CAGCCAGGTGGG.....CGCACCGCGGGCGCAACCCACCCGACGAGG 67
||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
478 GlnAlaGlyGlyProArgGlyArgAlaArgGlyProArgGlnGlnAl 494
: ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
494 aArgArgHisGlyProGlnArgArgGly.....Pro.Pro 507
118 CTGCTGTGTGATCAGCGCCAGCGCGCGCGCGCGCGCTGCTGTCG 159
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
508 GlnAlaGlyGluGlyProGlyAspAlaThrLeuValLeuGlyLeuG 524
160 .....CGCGCGAGTCGCTACTGCGC 181
524 yThrSerGlyGluGlnArgAlaAspGlnSerGlnThrLeuProAla 541
182 CAGTGGTGTGTCGACCG.CACCGCGGTGATGTCACGTGATCGTAA 230
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
541 euAlaGlyAlaProThrAlaHisAlaHisAlaValProGlyPro 557
231 GCGGTGTCGCCCGCGGTGTGTCGCGC.....GCGTGTTCGCGGAT 274
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
558 AlaAlaThrLeuGlyGlyArgGlyArgGlySerTrpArg..... 572
275 CGGTGCGGTGGCGCGCTCGGTGGTGGCGGAGCGATGGCGCGGT 324
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
573 .....GlyGlyArgGlyGlyGlyAlaGlyAlaSerGlyGlyAla 587
325 TCGCAATCGG.....CGGTCCACCGCGCGTGTGTCGCGCC 365
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
587 rGlyGlyArgGlyArgGlyArgGlyArgGlyArgGlySerGlyLeu 603
366 GGCACCGCTCGC...GCAGGACGCTGAAGAAGACGACGACGACG 412
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
604 GlyThrArgGluAspAlaGlySerProSerAlaArgArgGlyGln 620
413 ACRAAGAGGACGCTGTGACCTCCGTAATGACAACAGACTTCCG 462
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
620 gArgArgGlyHisGlyProProAlaAlaGlyAlaAlaGlnValSer 637
463 ACCCGGCGCGAAGACTTGCACACTTTTGGCGAGGAGGTAAAGAG 509
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
637 rGlyArgArgAlaArgGlyGlnArgThrGlyGluGluAlaGlnAsp 652
```

seq\_name: SwissProt\_40:KLF2\_HUMAN

seq\_documentation\_block:

ID KLF2\_HUMAN STANDARD; PRT; 355 AA.  
AC Q9Y5W3; Q9UKR6; Q9UJS5;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 01-MAR-2002 (Rel. 41, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Kruppel-like factor 2 (Lung kruppel-like factor).  
GN KLF2 OR LKLF. (Human).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Lung;  
RX MEDLINE=99231781; PubMed=10217429;  
RA Kozyrev S.V., Hansen L.L., Poltarus A.B., Dominsky D.A.,  
RA Kisselev L.L.;  
RT "Structure of the human CpG-island-containing lung Kruppel-like factor  
RT (LKLF) gene and its location in chromosome 19p13.11-13 locus.";  
RL FEBS Lett. 448:149-152(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Lung;  
RX MEDLINE=99389728; PubMed=10458913;  
RA Wani M.A., Konright M.D., Jeffries S., Hughes M.J., Lingrel J.B.;  
RT "cDNA isolation, genomic structure, regulation, and chromosomal  
RT localization of human lung kruppel-like factor.";  
RL Genomics 60:78-86(1999).  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Lee H.J., Kim M.K., Kim Y.H., Seo J.M., Lee H.M., Chung H.J.,  
RA Sohn M.Y., Hwang S.Y., Im S.U., Jung E.J., Kim J.C.;  
RT "A catalogue of genes in the human dermal papilla cells as identified  
RT by expressed sequence tags.";  
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: BINDS TO THE CACCC BOX IN THE BETA-GLOBIN GENE PROMOTER  
CC AND ACTIVATES TRANSCRIPTION (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).  
CC -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-  
CC FINGER PROTEINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
DR EMBL; AF123344; AAD25076.1; -;  
DR EMBL; AF134053; AAD55891.1; -;  
DR EMBL; AF205849; AAP13295.1; -;  
DR HSSP; P08047; 1SP2.  
DR MIM; 602016; -;  
DR InterPro; IPR000822; Znf-C2H2.  
DR Pfam; PF00056; Znf-C2H2; 3.  
DR SMART; SM00355; Znf-C2H2; 3.  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 3.  
DR PROSITE; PS0157; ZINC\_FINGER\_C2H2\_2; 3.  
KW Transcription regulation; Activator; Zinc-finger; Metal-binding;  
KW DNA-binding; Nuclear protein; Repeat.  
FT DOMAIN 62 71 POLY-PRO.  
FT DOMAIN 130 135 POLY-GLY.  
FT DOMAIN 167 171 POLY-PRO.  
FT DOMAIN 225 231 POLY-ALA.  
FT DOMAIN 272 354 ZINC FINGERS.  
FT 2N\_FING 272 296 C2H2-TYPE.  
FT 2N\_FING 302 326 C2H2-TYPE.  
FT 2N\_FING 332 354 C2H2-TYPE.  
FT CONFLICT 43 43 S -> N (IN REF. 2).  
FT CONFLICT 104 104 L -> P (IN REF. 1).  
FT CONFLICT 175 175 P -> S (IN REF. 2).  
FT CONFLICT 184 184 L -> M (IN REF. 2).  
SQ SEQUENCE 355 AA; 37419 MW; D5849C831D676AE1 CRC64;

alignment\_scores:

Quality: 125.00 Length: 217  
Ratio: 1.289 Gaps: 13  
Percent Similarity: 44.700 Percent Identity: 29.032

alignment\_block:

US-09-462-480-2/rev x KLF2\_HUMAN











oriented open reading frame: characterization of their promoter and enhancer regions.";  
Virology 179:365-377(1990).  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

EMBL; M34651; AAA47471.1; -  
PIR; B45344; B45344.  
FT DOMAIN 112 117 POLY-THR.  
FT DOMAIN 179 1733 GLY-RICH.  
FT DOMAIN 192 196 POLY-SER.  
FT DOMAIN 271 298 POLY-PRO.  
FT DOMAIN 304 308 POLY-ARG.  
FT DOMAIN 883 889 POLY-GLY.  
FT DOMAIN 1398 1405 POLY-GLY.  
SEQUENCE 1733 AA; 172166 MW; 0C8CD8BB475BB5E2 CRC64;

## alignment\_scores:

Quality:	119.50	Length:	231
Ratio:	1.160	Gaps:	12
Percent Similarity:	44.589	Percent Identity:	28.571

## alignment\_block:

US-09-462-480-2 x VNUA\_PRVKA ..

Align seg 1/1 to: VNUA\_PRVKA from: 1 to: 1733

31	GTGGCGCGCACCGG	.....	GGCGG	50
840	ValGlyGluGlyArgLeuGlyGlyProArgValGlyLeuAlaG	.....	GGCGG	856
51	CAACCCAGCCGACGAGGAGCGCGGCTGCTCGGCGCACGATC	.....	GGCGG	100
856	YargAspAlaGluAlaValGlyArgGlyValLeuGlyHisGlyP	.....	GGCGG	873
101	CG...CTGTGCAACATCCGCTG.....GCTGTGTGATCAGCG	.....	GGCGG	135
873	roGluArgAlaProGluProValValLeuGlyGlyGlyGlyGly	.....	GGCGG	889
136	CCCAGCGCGCGCGGCGCTG.....	.....	GGCGG	156
890	GlnGlnArgGlySerGlyValArgSerGlyProGluSerGlyAlaAl	.....	GGCGG	906
157	.....	.....	GGCGG	157
906	aLeuAlaProGlyProProValLeuPheValValAlaValAlaVal	.....	GGCGG	923
158	TGCGCGCGGAG.....TCGCTACCT	.....	GGCGG	177
923	alProAlaGluGlyArgAlaGlyGluProLeuValLeuAlaValPro	.....	GGCGG	939
178	GGCGCGGTGGTGTGACCCGCGGCTGATGCTCAGCTCATCGA	.....	GGCGG	227
940	GlyAlaAlaGly.....ProGlyArgAlaAlaLeuLeuLe	.....	GGCGG	951
228	AAAGCCGGTTCGCCCTCGGTGATGCCGCGGTGTTCGCCGATCGTGG	.....	GGCGG	277
951	uAlaProLeuGlyArgTrpValArg...AlaGlyGlyGlyAlaGlyV	.....	GGCGG	967
278	TGACGGGTGGCGCGCTCGGTGGTGGTGGCGGAGCG.....	.....	GGCGG	312
967	alAlaGlyGlyAlaGlyGluAlaGlyLeuGlyAlaGlyAlaGlyLeuGly	.....	GGCGG	983
313	ATGGCCAGAGGTTCGCAATCCGGCGGCTCCAC.CAGCCCGGCTGTGGTCG	.....	GGCGG	361
984	AlaGlyAlaGlyLeuGlyAlaGlyGlyAlaGlyGlyProGlyAlaGlyG	.....	GGCGG	1000

362	CGCGCGCACCGCTCGCGAGGAGCGTGAAGAAGACGA.....	398
1000	uAlaGly.....GlyGlyAlaArgArgArgArgArgTrpA	1014
399	.....CGAGGACGACTGGGACGAGA	419
1014	spAspGluAlaGlyLeuLeuGlyProGluArgGlyGlnAlaGlyArgGly	1030
420	GGAGGACTGGTGCCTCCGTAATGACACAGACTTCCC.....	GGCCA 463
1031	LeuArgGlyProGlyProArgGlyGlyLeuGlyGluProGlyArgGlyHI	1047
464	CCCGCGCGGAGGAGACTTCCCAACATTTTGGCGAGGAGGTAATA	506
1047	sValGlyArg.....GlyGluGluGlyArg	1055

seq\_name: SwissProt\_40:EXLP\_TOBAC

## seq\_documentation\_block:

ID EXLP\_TOBAC STANDARD; PRT; 426 AA.

AC Q03211;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE Pistil-specific extensin-like protein precursor (PELP).  
OS Nicotiana tabacum (Common tobacco).  
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.  
OX NCBI\_TaxID=4097;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV, PETITE HAVANA; TISSUE=Pistil;  
RX MEDLINE=93005740; PubMed=1392607;  
RA Goldman M.H., Pezzotti M., Seurinck J., Mariani C.;  
RT "Developmental expression of tobacco pistil-specific genes encoding novel extensin-like proteins.";  
RL Plant Cell 4:1041-1051(1992).  
CC -!- TISSUE SPECIFICITY: PISTIL (STIGMA AND STYLE TISSUE).  
CC -!- DEVELOPMENTAL STAGE: EXPRESSION BEGINS IN FLORAL BUDS AFTER PISTIL DIFFERENTIATION AND LEVELS GRADUALLY INCREASE DURING FLOWER DEVELOPMENT TOWARD ANTHESIS. LEVELS GRADUALLY DECREASE AFTER POLLINATION AND ARE ABSENT BY THE SIXTH DAY AFTER POLLINATION.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; Z14019; CAA78397.1; -  
CC PIR; JQ1696; JQ1696.  
CC InterPro; IPR000419; Pollen\_Ole\_e\_I.  
CC Pfam; PF01190; Pollen\_Ole\_e\_I; 1.  
CC Structural protein; Signal; Repeat; Glycoprotein.  
CC SIGNAL 1 23  
CC CHAIN 24 426 PISTIL-SPECIFIC EXTENSIN-LIKE PROTEIN.  
CC DOMAIN 69 182 4 X 5 AA REPEATS OF S-P(4).  
CC REPEAT 69 73 1.  
CC REPEAT 76 80 2.  
CC REPEAT 83 87 3.  
CC REPEAT 178 182 4.  
CC CARBOHYD 310 310 N-LINKED (GLCNAC...) (POTENTIAL).  
CC SEQUENCE 426 AA; 44278 MW; 51A495C94017812 CRC64;

## alignment\_scores:

Quality:	119.00	Length:	175
Ratio:	1.451	Gaps:	9

Percent Similarity: 46.857 Percent Identity: 28.571

## alignment\_block:

US-09-462-480-2/rev x EXLP\_TOBAC ..

Align seg 1/1 to: EXLP\_TOBAC from: 1 to: 426

```
524 GCTGACACTTCTCTCTTTACCTTCCTCGCCAAATGTTGGCAAGCT 475
|||||
113 AlAGlyLeuProLeu.IleProAsnLeuProAspValProProIleG 129
474 TC.....CGGCGGGTGGCGGGAAGTGTG 449
129 lyGlyProProValAsnGlnProLysProSerProSerProLeu 145
448 TGTGATTACGGAGCTCACCAGTCTCTCTCTGTCCTCCAGTCGCTCCG 399
146 VallyProProProProSerProCysLysProSerProProAs 162
398 TCGTCTCTTCACTCTCGCTCGCGAGCGGTGCGCGCGACGACCGCG 349
162 pGlnSerAlaLysGlnProGlnProProAlaLys..... 175
348 GCTGTGAGCGCGCGGATTGGGAACCTTGGCCCATCGCTCCGGAACCA 299
176 .....GlnProSerProPro...ProProPro 183
298 CGGAGCGCGCGCACCTCGACGAGTCCGGCAACAGCGCGCGCATC 249
184 ProProVallyAlaProSerProSerProAlaLysGlnProProPr 200
248 ACCGAGGGGCAACGGCTTTTCGATCAGCTGAGACATCAGCGCGTGC 200
200 oPro.....ProProVallyAlaProSerProSerProAlaThrg 214
199 G.....GTCAACGACCCACTGCGCCAGTAGTACCTCGCGCGGAGC 156
214 InProProThryLysGlnProProProPro.....ProArgAla 226
155 AGGCGCGCGCGCTGGGCGCTGATCCACGACGCGGATGTTTGA 106
227 LysLysSerProLeuLeuProProProProProValAla.....TyrPr 241
105 CAGCGGACTGTGCGGAGCAGCCCATCTGCGCGCTTCCCTCGTGGCTG 56
241 oProValMethThrProSer...ProSerProAlaAlaGluProProIleI 257
55 GGTTCGCGCGCGGTGCGCGCC 33
257 leAlaProPheProSerProPro 264
```

seq\_name: SwissProt\_40:FMN\_CHICK

## seq\_documentation\_block:

```
ID FMN_CHICK STANDARD; PRT; 1213 AA.
AC Q05858;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Formin (Limb deformity protein).
GN LD.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WHITE LEHORN; TISSUE-Embryo;
RX MEDLINE=92112031; PubMed=1730407;
RA Trumpp A., Blundell P.A., de la Pompa J.L., Zeller R.;
RT "The chicken limb deformity gene encodes nuclear proteins expressed
in specific cell types during morphogenesis.";
```

```
Genes Dev. 6:14-28(1992).
-1- FUNCTION: IS IMPORTANT FOR MORPHOGENESIS OF LIMB AND KIDNEY AND
MAY BE INVOLVED IN DETERMINING DORSOVENTRAL NEURAL TUBE POLARITY
AND MOTOR NEURON INDUCTION. IT MAY ALSO HAVE A FUNCTION IN
DIFFERENTIATED CELLS OR BE INVOLVED IN MAINTAINING SPECIFIC
DIFFERENTIATED STATES.
-1- SUBCELLULAR LOCATION: Nuclear.
-1- ALTERNATIVE PRODUCTS: MANY DIFFERENT FORMS OF THIS PROTEIN ARE
PRODUCED BY ALTERNATIVE SPLICING OF THE LD GENE. A VARIATION
IN SPLICING IS SEEN AMONG DIFFERENT TISSUES AND DIFFERENT SIZE
TRANSCRIPTS EXIST WITHIN ANY ONE TISSUE.
-1- TISSUE SPECIFICITY: PRESENT IN THE ADULT BRAIN, KIDNEY,
BRAIN, HEART AND INTESTINE AND THROUGHOUT THE EMBRYO.
-1- DEVELOPMENTAL STAGE: IN THE DEVELOPING LIMB BUD, THE PROTEIN IS
EXPRESSED IN THE APICAL ECTODERMAL RIDGE AND THE MESENCHYMAL
COMPARTMENT. PREDOMINANTLY IN THE POSTERIOR REGION. DURING
KIDNEY MORPHOGENESIS, EXPRESSION IS INITIALLY RESTRICTED TO
THE EPITHELIAL COMPARTMENT OF THE PRONEPHROS AND MESONEPHROS.
-1- SIMILARITY: CONTAINS 1 FORMIN HOMOLGY 1 (FH1) DOMAIN.
-1- SIMILARITY: CONTAINS 1 FORMIN HOMOLGY 2 (FH2) DOMAIN.
-1- SIMILARITY: BELONGS TO THE FORMIN HOMOLGY FAMILY. CAPPUCCINO
SUBFAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; X62681; CAA44555.1; -.
DR InterPro; IPR003104; FH2.
DR InterPro; IPR001265; Formin.
DR Pfam; PF02181; FH2; 1.
DR PRINTS; PR00828; FORMIN.
DR SMART; SM00498; FH2; 1.
DR Nuclear protein; Developmental protein; Coiled coil;
KW Alternative splicing.
FT DOMAIN 428 450 COILED COIL (POTENTIAL).
FT DOMAIN 503 572 COILED COIL (POTENTIAL).
FT DOMAIN 652 751 FH1 (PRO-RICH).
FT DOMAIN 766 1171 FH2.
FT DOMAIN 1050 1125 COILED COIL (POTENTIAL).
SQ SEQUENCE 1213 AA; 135240 MW; ADE3EF0B3FB9D862 CRC64;

alignment_scores:
Quality: 118.00 Length: 107
Ratio: 2.511 Gaps: 5
Percent Similarity: 43.925 Percent Identity: 31.776

alignment_block:
US-09-462-480-2/rev x FMN_CHICK ..
Align seg 1/1 to: FMN_CHICK from: 1 to: 1213
353 CCCGGCTGTGGAGCGCGGATTCGGAACCTGGCCCATCGCTCCCGG 304
|||||
678 ProGlyLeuValProProPro.....ProLeuProThrGl 690
303 ACCACCGGAGCGCGCACCGCTGACCGAGCATCCGGCAACAGCGCGG 254
|||||
690 yProThrSerValThrProHisPheAlaPheGlyProProLeuProProG 707
253 GCATCACCAGGGGCAACCGCTTTTCGATCAGCTGAGACATCAGCGCG 204
|||||
707 InLeuSerGluGlyCysArgAspPhe..... 715
203 GTGCGGGTCAACGACCGCCAGCTGCGGAGTAGCGAGTCCGCGCGCAGCAG 154
|||||
716 ...GlnAlaProAlaProProAlaPro.....Pr 724
```







OM of: US-09-462-480-2 to: SPTREMBL\_19:\* out\_format: pfs  
Date: Jul 22, 2002 1:40 AM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:  
-MODEL=frame+n2p.model -DEV=xlh  
-O=/cgn2.1/USPTO.spool/US09462480/runat\_18072002.164119.19544/app\_query.fasta\_1.2850  
-DB=SPTREMBL\_19 -OFMT=fastan -SUFFIX=rspt -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blotsum62  
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09462480\_EGNI\_1\_133 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-462-480-2  
Query length: 524  
Database: SPTREMBL\_19:\*  
Database sequences: 562222  
Database length: 172994929  
Search time (sec): 143.350000

score_list:	Sequence	Std	Orig	ZScore	Escore	Len	Documentation
	sp_bacteriapi:O69738	713.00	952.62	2.8e-45	368	1	O69738 mycobacterium tuberculosis
	sp_bacteriapi:O33085	198.00	265.58	6.3e-07	302	1	O33085 mycobacterium leprae. hy
	sp_bacteria:Q9EY22	145.50	196.92	0.0051	251	1	Q9EY22 paracoccus pantotrophus.
	sp_human:Q9B037	143.50	191.27	0.0073	358	1	Q9B037 homo sapiens (human). si
	sp_bacteria:Q9XDH2	136.50	175.58	0.0257	763	1	Q9XDH2 mycobacterium tuberculosis
	sp_bacteria:Q9Y164	133.00	169.40	0.0474	913	1	Q9Y164 actinomycetes viscosus. si
	sp_virus:P89459	132.50	158.44	0.0565	3122	1	P89459 herpes simplex virus (t
	sp_plant:Q9FPQ6	131.50	171.56	0.0591	555	1	Q9FPQ6 chlamydomonas reinhardtii
	sp_invertebrate:Q20739	129.50	173.92	0.0798	304	1	Q20739 caenorhabditis elegans
	sp_human:Q9UKR6	129.00	171.95	0.0879	355	1	Q9UKR6 homo sapiens (human). ki
	sp_mammal:Q9N1P0	128.50	164.30	0.1018	818	1	Q9N1P0 bos taurus (bovine). sub
	sp_invertebrate:Q9NHW4	128.50	155.84	0.1096	2249	1	Q9NHW4 nephila clavipes (orb
	sp_invertebrate:O44358	128.00	163.11	0.1114	871	1	O44358 nephila clavipes (orb
	sp_rodent:Q9WUE8	127.50	155.13	0.1294	2087	1	Q9WUE8 rattus norvegicus (rat)
	sp_rodent:Q9WU13	127.50	154.85	0.1297	2158	1	Q9WU13 rattus norvegicus (rat)
	sp_rodent:Q9WV48	127.50	154.81	0.1298	2167	1	Q9WV48 rattus norvegicus (rat)
	sp_rodent:Q9ES29	127.00	162.38	0.1316	810	1	Q9ES29 mus musculus (mouse). el
	sp_plant:Q9SX31	125.50	161.50	0.1685	708	1	Q9SX31 arabidopsis thaliana (m
	sp_bacteria:Q9Y164	125.50	159.37	0.1716	913	1	Q9Y164 actinomycetes viscosus. si
	sp_virus:Q69340	125.50	152.99	0.1815	1958	1	Q69340 pseudorabies virus. orb
	sp_human:Q9U0J5	125.00	166.60	0.1745	355	1	Q9U0J5 homo sapiens (human). ki
	sp_bacteria:Q9ZBP2	125.00	164.51	0.1778	456	1	Q9ZBP2 streptomyces coelicolor.
	sp_bacteria:Q9F2N5	125.00	158.92	0.1867	889	1	Q9F2N5 streptomyces coelicolor.
	sp_plant:Q9ZRH8	124.50	164.90	0.1919	402	1	Q9ZRH8 oryza sativa (rice). ear
	sp_invertebrate:Q9VZC2	124.50	164.03	0.1934	446	1	Q9VZC2 drosophila melanogaste
	sp_bacteria:Q68872	124.50	162.40	0.1961	542	1	Q68872 myxococcus xanthus. hyp
	sp_bacteriapi:Q33268	124.50	158.32	0.2032	882	1	Q33268 mycobacterium tuberculosis
	sp_bacteria:Q69995	124.00	159.08	0.2187	744	1	Q69995 streptomyces coelicolor.
	sp_bacteriapi:Q06556	123.00	167.67	0.2380	227	1	Q06556 mycobacterium tuberculosis
	sp_virus:Q69270	123.00	161.60	0.2510	469	1	Q69270 equine herpesvirus 1. ir
	sp_virus:Q9IPQ8	123.00	159.17	0.2552	588	1	Q9IPQ8 cynomolgus epstein-barr
	sp_mammal:Q9N1P1	123.00	159.12	0.2565	631	1	Q9N1P1 bos taurus (bovine). sub
	sp_human:Q9Y2J6	122.50	156.76	0.2836	772	1	Q9Y2J6 homo sapiens (bovine). sub
	sp_invertebrate:Q9V094	122.50	153.13	0.2928	1192	1	Q9V094 drosophila melanogast
	sp_invertebrate:Q9G0I6	122.50	153.12	0.2928	1193	1	Q9G0I6 drosophila melanogast
	sp_invertebrate:Q9VX54	122.00	159.02	0.3012	544	1	Q9VX54 drosophila melanogaste
	sp_human:Q9G508	122.00	152.51	0.3189	1184	1	Q9G508 homo sapiens (human). s
	sp_invertebrate:Q9NHW2	122.00	148.63	0.3299	1884	1	Q9NHW2 nephila madagascari
	sp_rodent:Q88493	122.00	145.75	0.3383	2657	1	Q88493 mus musculus (mouse). t
	sp_invertebrate:Q9ET53	121.50	162.71	0.3159	323	1	Q9ET53 mus musculus (mouse). 9d

sp_virus:Q9Q5K9	-	121.50	157.42	0.3309	608	1	Q9Q5K9 herpesvirus papio. nt
sp_rodent:Q99K31	-	121.50	157.31	0.3312	616	1	Q99K31 mus musculus (mouse).
sp_plant:Q9LD34	-	121.50	156.97	0.3322	642	1	Q9LD34 cryptosporidium cohnii
sp_human:Q9P2P0	-	121.50	152.06	0.3468	1154	1	Q9P2P0 homo sapiens (human)
sp_bacteria:Q9RRK9	-	121.50	150.85	0.3505	1334	1	Q9RRK9 streptomyces coellic
seq_name: sp_bacteriapi:O69738							
seq_documentation_block:							
ID	O69738				368	AA.	
AC	O69738:	PRELIMINARY;	PRT;				
DT	01-AUG-1998	(TREMBlrel. 07, Created)					
DT	01-AUG-1998	(TREMBlrel. 07, Last sequence update)					
DT	01-DEC-2001	(TREMBlrel. 19, Last annotation update)					
DE	PPE-FAMILY	PROTEIN.					
GN	RV3873	OR MT027.08.					
OS	Mycobacterium tuberculosis.						
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;						
OC	Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.						
OX	NCBI_TaxID=1773;						
RN	[1]						
RP	SEQUENCE FROM N.A.						
RC	STRAIN=H37RV;						
RX	MEDLINE=98295987; PubMed=9634230;						
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,						
RA	Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,						
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,						
RA	Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,						
RA	Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,						
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,						
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R.,						
RA	Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;						
RT	"Deciphering the biology of Mycobacterium tuberculosis from the						
RT	complete genome sequence."						
RL	Nature 393:537-544(1998).						
DR	EMBL; AL022120; CAA17965.1; -.						
DR	Tuberculist; RV3873; -.						
DR	InterPro; IPR000030; PPE.						
DR	Pfam; PF00823; PPE; 1.						
KW	Complete proteome.						
SQ	SEQUENCE 368 AA; 37330 MW; D78F44095F658CA2 CRC64;						
alignment_scores:							
	Quality:	713.00	Length:	143			
	Ratio:	5.057	Gaps:	0			
	Percent Similarity:	98.601	Percent Identity:	96.503			
alignment_block:							
US-09-462-480-2 x O69738 ..							
Align seg 1/1 to: O69738 from: 1 to: 368							
1	CTGCAGAGTGACGTGGTGTTCAGCCAGTGGCGGCACCGCGCGGG	50					
226	LeuGlnGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGly	242					
51	CAACCCAGCGACGAGCAAGCCGCGAGATGGCGCTCTCGGCACCA	100					
242	YASNPrOAlaAspIuGluAlaAlaGlnMetGlyLeuLeuGlyThrSer	259					
101	CGTGTGTCGAACCATCCCGTGGCTGGTGGATCAGGCCCGCCAGCG	150					
259	roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla	275					
151	GGCGTGTGTCGGCGGAGTGCCTACCTGGCGCAGGTGGTGGTGGAC	200					
276	GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThr	292					
201	CAGCGCCCTGATGTCCTCAGCTGATCGAAAGCGGTTGCCCTCGCT	250					
292	gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerVal	309					



```

366 CGGCG.....CGACGACCCGGGCTGGTGGAGC...CGCGGGATT 329
    ::::  |||||  |||  |||  |||||:::
113 laSerArgLysAsnArgProLysProTrpArgSerArgArgVal 129
    ::::  |||||  |||  |||  |||||:::
328 CGAACCTGGC.....CCATCGCTCCCGGACCCACCGGA 294
    ::::  |||||  |||||  |||||  |||||
130 SerLysProSerArgArgSerArgProSerSerLeuProArgProHi 146
    ::::  |||||  |||||  |||||  |||||
293 CGCGCGCACCGCCACCGAGGATCCGGCAACACCGCGGCATCACCGA 244
    ::::  |||||  |||  |||:  |||:  |||:  |||:  |||:  |||:
146 sArgArgThrProArgArgHisArgMetArgArgProGlySer.... 161
243 GGGGGCAACGGCTTTTCGATCAGCTGAGACATCAGCGCGCTCGGGTCA 194
    |||  ::::  |||  ::::  |||  ::::  |||  ::::  |||  ::::
162 .....ArgLeuProSerLeuProSerAlaArgArgPro 172
193 ACGACCCACCTGCCGAGTAGGACTCCGGCGCAGCAGCGCCCGCGCC 144
    |||  ::::  |||  ::::  |||  ::::  |||  ::::  |||  ::::
173 ThrProArgLeuLeuAsnProArgProThrProArgLysProArgPr 189
143 GCGCTGGGGCTGATCCACCGACCGGATGTTTCGACACGCGACTGGT 94
    ||||  ::::  |||  |||  |||||  |||||
189 oArg..AsnProLysPro.....SerGlyGlnAr 198
93 GCCGAGCAGGCCCATC.....TGGCGGCTTCTCGTCGGCTG 56
    ||||  ::::  |||  |||  |||  |||  |||  |||  |||  |||
198 gProLysArgProLeuHisProLysProCysArgArgProSerProSerP 215
55 GGTTCGGCGCGCGGTGCGGCC 33
    |||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||
215 roProProSerProLeuProPro 222
seq_name: sp_human:Q9BU37
seq_documentation_block:
ID Q9BU37 PRELIMINARY; PRT; 358 AA.
AC Q9BU37;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SIMILAR TO WISKOTT-ALDRICH SYNDROME PROTEIN INTERACTING PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MELANOMA.;
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC002914; AAH02914.1; -.
DR InterPro; IPR002965; P_rich_extensn.
DR InterPro; IPR003124; WH2.
DR Pfam; PF02205; WH2; 1.
DR PRINTS; PR01217; PRICHEXTENSIN.
DR SMART; SM00246; WH2; 1.
SQ SEQUENCE 358 AA; 36464 MW; D008B60E60EE94EA CRC64;

```

```

alignment_scores:
  Quality: 143.50      Length: 187
  Ratio: 1.966
Percent Similarity: 39.037 Percent Identity: 29.947

alignment_block:
US-09-462-480-2/rev x Q9BU37 ..
Align seg 1/1 to: Q9BU37 from: 1 to: 358

508 TCTTTACTCTCCGCCAAATGTGGCAAGTCTCCGGCCCGCGGTGCC 459
    |||  |||||:  |||  :::  |||  |||||
142 SerProProSerGlyProGlyArgPheProValProSerProGlyHisAr 158

```

```

458 GGAAGTCTGTGTTCATTACGGGAGCTCACCAGTCGTCTCTTCGTCCCA 409
    ::::  |||||  |||  |||  |||||  |||
158 gSerGly.....Prop 162
408 GTCGTCTCTGTCGTCTTTCACGCTCTCGCGAGCGGTGCCGGCGGA 359
    ::::  |||||  |||  |||  |||||  |||
162 roGluProGlnArg.....AsnArgMetProProProArg 173
358 CCACAGCCGGGCTGGTGGAGCCGGGATTGGGAACCTGCCCATCGCT 309
    |||||  |||  |||  |||  |||||  |||
174 ProAspValGlySerLysProAspSerIleProProProValProSerTh 190
308 CCCGAGACCA..... 299
    |||  |||
190 rProArgProIleGlnSerSerLeuHisAsnArgLysSerProProValP 207
298 .....CCGAGCGGGCCACCGGTCCACGAGATCCGGCAACAGCGC 257
    |||  ::::  |||||  |||||  |||||  |||
207 roGlyGlyProArgGlnProSerProGlyProThrProProProPro 223
256 CCGCATCACCGAGGGGCAA.....CCGCGC 231
    ::::  ||||  ||||  ||||  ||||  ||||  ||||  ||||
224 ValArgAspProProGlyArgSerGlyProLeuProProProProPro 240
230 TTTTCGATCAGCTGAGCATCAGCGGCGTGGGGTCAACACCGCACCTGC 181
    ::::  ||||  ||||  ||||  ||||  ||||  ||||  |||
240 lSerArgAsnGlySerThrSerArgAlaLeuProAlaThr..ProGlnLe 256
180 GCCA.....GGTAGCGACTCGCGCGCAGCAGCGCGCGCGCGC 140
    |||  |||  |||||  |||||  |||||  |||  |||  |
256 uProSerArgSerGlyValAspSerProArgSerGlyProArgProProL 273
139 TGGGGCTGATCCACCGACCGGATGTTTCGACACGCGACTGGTGGCG 90
    ||  |||||  |||
273 euProProAspArgPro..... 278
89 AGCAGGCCCATCTGCGCGCTTCTCGTGGTGGTTCGCGCGCGCGCGT 40
    |||||  |||||  |||||  |||||  |||||
279 .....SerAlaGlyAlaProProProPro 286
39 GCCGCGCCACC 30
    |||||  ::::
286 oProProSer 289
seq_name: sp_bacteria:Q9XDH2
seq_documentation_block:
ID Q9XDH2 PRELIMINARY; PRT; 763 AA.
AC Q9XDH2;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PROLINE-RICH MUCIN HOMOLOG.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=20090472; PubMed=10627046;
RA Espitia C., Lacleite J.P., Mondragon-Palomino M., Amador A.,
RA Campuzano J., Martens A., Singh M., Cicero R., Zhang Y., Moreno C.;
RT "The PE-PGRS glycine-rich proteins of Mycobacterium tuberculosis: a
RL Microbiology 145:3487-3495(1999).
DR EMBL; AF071081; AAD41594.1; -.
DR InterPro; IPR002951; Atrophin.
DR InterPro; IPR003882; Pistil_extensin.
DR InterPro; IPR002965; P_rich_extensn.
DR PRINTS; PR01222; ATROPHIN.
DR PRINTS; PR01217; PRICHEXTENSIN.
DR PRINTS; PR01218; PSTLEXTENSIN.

```

```
SQ SEQUENCE 763 AA; 75035 MW; 39168EC45A5916F8 CRC64;

alignment_scores:
  Quality: 136.50      Length: 166
  Ratio: 1.706         Gaps: 10
  Percent Similarity: 48.193      Percent Identity: 34.337

alignment_block:
US-09-462-480-2/rev x Q9XDH2 ..
  Align seg 1/1 to: Q9XDH2 from: 1 to: 763

505 TTACTTCTCCGCCCAAAATGTGGCAAGTCTTCGGCGCCGGTGGCCGGG 456
||||| ||||| :
538 LeuProProSerProProAla.....ProAsnSerProPr 549
455 AAGTCTGTGTCATTCAGGGAGTCACAGTCGCTCTTCGTCACGATC 406
: : |||
549 oAla.....ProProAlaProProT 556
405 GTCCTCGTCGCTCTTTCACGCTCTCGCGGAGCGGTGCGCGCGACCA 356
||| :||| ||| :||| :||| :|||
556 hrProProLysLeuLeuSerAlaAsnProProCysProProValProPro 572
355 GACCCGGGCTGGTGGAGCGCGGATTGCGAACCCCTGGC..... 317
||| :||| ||| ||| :|||
573 AlaPro.....AsnArgProProAlaProProAlaProProAlaPr 586
316 .CCATCGCTCCGACGACCGGAGCGCGCCACCGTCACCGAGATCC 268
||||| ||||| :||| :||| :|||
586 oProGluLeuProAlaProProAsp.....ProProThrProProVala 601
267 GGCAACAGCCGCGGCATCACGAGGGGCAACCGGCTTTTCGATCAGCT 218
||| :||| :||| :||| :|||
601 laAsnSerProProAlaProProAla.....ProProAla 612
217 GAGCATCATCAGCGGTGGTGGTCAACGACCCACCGTGGCGAGTAGGAC 168
||||| :||| :||| :||| :|||
613 ProProSerAla..LeuProPheValAsnProProAlaProProThrPro 628
167 TCCGGG...CGCAGCAGCGCGG...CCGCGCTGGGCTGTATCCACC 124
||||| :||| :||| :||| :|||
629 AlaAlaProLysSerArgProAlaLeuProAlaAlaProProAlaProPr 645
123 AGCCAGCGGATGTTGACACGCGACTGGTGGCGAGAGCGCCATCTGGG 74
||| :||| :||| :||| :|||
645 oAlaProValArgAlaThrProProProAlaProProAlaProp 662
73 CGGTTCTCTCGTGGTGGTGGTGGG.....CGCGCGGTGCGCGCCC 33
||| :||| :||| :||| :|||
662 roAlaProAsnSerMetAlaLeuProProAlaProProAspProPro 677

seq_name: sp_bacteria:Q59164

seq_documentation_block:
ID Q59164 PRELIMINARY; PRT; 913 AA.
AC Q59164;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE STALIDASE (EC 3.2.1.18).
GN NANH.
OS Actinomyces viscosus.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Actinomycineae; Actinomycetaceae; Actinomycetes.
OX NCBI_TaxID=1656;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM43798;
RA Henningsen M.;
RL Submitted (SEP-1991) to the EMBL/GenBank/DBJ databases.
[2]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN=DSM43798;
RX MEDLINE=92162190; PubMed=1789931;
RA Henningsen M.; Roggentin P.; Schauer E.R.;
RT "Cloning sequencing and expression of the sialidase gene from
RT Actinomyces viscosus DSM 43798.";
RL Biol. Chem. Hoppe-Seyler 372:1065-1072(1991).
DR EMBL; X62276; CAA44166.1; -.
DR HSSP; Q02834; LEUR.
DR InterPro; IPR002860; BNR.
DR Pfam; PF02012; BNR; 5.
DR Hydrolase; Glycosidase.
KW SEQUENCE 913 AA; 96216 MW; AD22CF56706FF373 CRC64;

alignment_scores:
  Quality: 133.00      Length: 188
  Ratio: 1.529         Gaps: 9
  Percent Similarity: 46.277      Percent Identity: 29.787

alignment_block:
US-09-462-480-2/rev x Q59164 ..
  Align seg 1/1 to: Q59164 from: 1 to: 913

472 CGGCCCGGTGGCGGGAAGTCTGTTCATTACGGAGCTCACCAGTCG 423
||||| ||| ||||| ||||| :
686 ArgProArgArgProArgArgAlaLeuSerProArgArgHisArgHisH 702
422 TCCCTCTCGTCCAGTCGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 373
||| ||||| ||||| ||| ||||| |||||
702 sProProArgProSerArgAlaLeuArg.....ProSerArgA 715
372 CGGTGCGCGCGG.....CGACCAGACCCGCGGTGGTGAGC 338
||| :||| :||| :||| :|||
715 laGlyProGlyAlaGlyAlaHisAspArgSerGluHisGlyAlaHisThr 731
337 CGCGCGATTGCGAACCTGCGTCGTCGCGGACCCACCGGAGCGCG 288
||||| :||| :||| :||| :|||
732 GlySerCysAlaGlnSerAlaProGluInThrAspGlyProThrAlaAl 748
287 CCACCGTCCAGCAGATCCGCGCAACCGCGCGGATCACCAGGGGGC 238
||||| :||| :||| :||| :|||
748 a...ProAlaProGluThrSerSerAlaProAlaAlaGluProThrGln 764
237 AACCGGCTTTTCGA.....TCAGCTCAG 215
||||| :||| :||| :||| :|||
764 laProThrValAlaProSerValGluProThrGlnAlaProGlyAlaGln 780
214 ACATCAGCGCG.....TGGGGTCAACGACCCACCTGCGCCAGGTAGC 171
||| :||| :||| :||| :|||
781 ProSerSerAlaProLysProGlyAlaThrGlyArgAlaProSerValva 797
170 GACTCGCGCGCCAGCGC..... 152
: :||| :||| :||| :||| :|||
797 laSnProLysAlaThrGlyAlaAlaThrGluProGlyThrProSerSers 814
151 .....CGCGCGCGCGCTGGGCG 134
814 erAlaSerProAlaProSerArgAsnAlaAlaProThrProLysProGly 830
133 CTG.....ATCCACCGACCGGATGTTTCACAGCGGACTG 96
: :||| :||| :||| :||| :|||
831 MetGluProAspGluIleAspArgProSerAspGlyThrMetAlaGlnPr 847
95 GTGCCGAGCAGCGCCATCTGCGCGGCTTCCTGCTCGGCTGGGTGCCGCC 46
||| :||| :||| :||| :|||
847 oThrGlyAla...ProAlaArgArgValProArgArg.....ArgA 860
45 GCGGTCGCGGCCCA 32
||||| |||||
860 rgArgArgArgPro 864
```



```
DR PRINTS; PRO1217; PRICHEXTENSIN.
DR PRINTS; PRO1218; PSTLXTENSIN.
SQ SEQUENCE 555 AA; 54219 MW; 6A584A90465502F5 CRC64;

alignment_scores:
  Quality: 131.50      Length: 159
  Ratio: 1.906         Gaps: 10
  Percent Similarity: 43.396 Percent Identity: 33.962

alignment_block:
US-09-462-480-2/rev x Q9FPQ6 ..
Align seg 1/1 to: Q9FPQ6 from: 1 to: 555

502 CTTCTCTCGCCAAATGTTGCAAGTCTTCGGCGCGGTGGCGGGAAG 453
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
63 ProAlaProProSer.....ProGlyProProSer... 72
452 TCTGTGTTCATTACGGGAGCTCACAGTCGTCCTCTTCGTCCTCCAGTCGTC 403
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
73 .....ProAlaProProSerProp 79
402 CTCGTGCTTCTTCACGCTCTCGCGGAGCGGTGCGCGCGCGACAC 353
||:|||||:|||||:|||||:|||||:|||||:|||||:
79 ro.....SerProAlaProProSerProAlaProProSer 90
352 CCGGCTGGTGAGCGCGGATTGCGAACCTCGGCCCATCGCTCCGGA 303
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
91 ProAlaProProSerProAlaProProSerProAlaProProSerProAl 107
302 CCCACCGAGCGCGGCCACCGCTCACCGACGATCCGGCAACAGCGCGG 253
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
107 aProProSerProAlaProProSerProAla.....ProProS 120
252 CATCAGCGAGGGGCAACCGCTTTCGATCAGTCAGATCAGATCAGCGG 203
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
120 erProProSerProAlaProPro..SerProSer..... 130
202 TCGGGTCAACGACCACTCGCGCAGGTAGCGACTCCGCGCGCAGCAGG 153
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
131 .....ProProAlaProProSerProProSer... 141
152 CCGCGCGCGCGCTGGGGCTGATCCA...CCAGCCAGCGGATGGTTCA 106
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
142 ProAlaProProLeuProProSerProAlaProProSer..... 154
105 CAGCGGACTGTCGCGAGCAGGCCCATCTCGCGGCTTCCTCGTGGCTG 56
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
155 .....ProSerProProAlaProProSerProSer..... 164
55 GGTTCGCGCGCGCGTGGCGCCGACC 30
165 .....ProProValProProSer 170

seq_name: sp_invertebrate:Q20739

seq_documentation_block:
ID Q20739 PRELIMINARY; PRT; 304 AA.
AC Q20739;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE F54B11.2 PROTEIN.
GN F54B11.2
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RA Swinburne J.;
RS SEQUENCE FROM N.A.
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
```

```
RN [2] SEQUENCE FROM N.A.
RX MEDLINE-99069613; PubMed-9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; Z70208; CAA94136.1; -.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR002486; Col_cuticle_N.
DR Pfam; PF01391; Collagen; 2.
DR Pfam; PF01484; Col_cuticle_N; 1.
SQ SEQUENCE 304 AA; 28767 MW; 3EF0C84088428F35 CRC64;

alignment_scores:
  Quality: 129.50      Length: 175
  Ratio: 1.455         Gaps: 12
  Percent Similarity: 50.857 Percent Identity: 33.143

alignment_block:
US-09-462-480-2/rev x Q20739 ..
Align seg 1/1 to: Q20739 from: 1 to: 304

483 GCAGAGTCTTCGGCGCGGTGGCGGAGTCTGTTCTATTACGGG... 436
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
95 GlyAlaSerSerGly...GlyGlnCysGluGlyCysCysAsnProGlyPr 110
435 .....AGTCACCAAGTCGTCCTTCGTCCTCCAGTCGTCCTCG 399
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
110 oProGlyValAlaGlyAsnPro.GlyLysProGlyLysProGlyLysPro 126
398 TCCTCTTCTTCACGCTCTCGCGGAGCGGTGCGCGCGCACACCGG 349
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
127 GlyAlaProGlyAsnProGlyAlaProGlyLysGlyAlaAlaValProCy 143
348 GCTGCTGGAG...CCGCGGATTGCGAACCTCGGCCATCGCTCCCGGAC 302
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
143 sGluAlaLysThrProProCysLysProCysPro.....AlaGlyP 158
301 CCACCGAGCGCGCCACCGCTCACCGACGATCCGCGGCAACAGCGCGG 252
||:|||||:|||||:|||||:|||||:|||||:|||||:
158 ropProGly.....ProProGlyProAspGlyProAlaGlyProAlaGly 172
251 ATCACCAGGGGGCAACCGGCTTTTCGATCAGTCAGACATCAGCGGCGT 202
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
173 ...ProAspGlyGluAlaGlySerProAlaAlaProSerProProGly.. 187
201 GCGGTCACGACCCACCTCGCCA.....GGTA 173
|||||:|||||:|||||:|||||:|||||:
188 .....ProProGlyProSerGlyProAlaGlyProAlaGlyA 200
172 GCGACTCCGCGCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTGC 123
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
200 snAspGlyAlaAlaGlyThrPro.....GlyProAspGlyPro 212
122 GCCAGCGGATGTTTCGACAGCGGACTGTCGCGGAGCAGCGCCATCTGCG 73
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
213 AlaGly.....GluSerThrTyProGluProAl 222
72 GGCTTCCTCGTCGGCTGGGTGCGCGCGCGCGCGCGCGCGCGCGCGTGA 23
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
222 aAlaProGlyProAlaGlyProProGlyProAlaGlyProProGlyPro 238
22 ACACGAGCGTCACTGCTGCA 2
|||||:|||||:|||||:|||||:|||||:
239 AspGlyAlaSerProThrAla 245

seq_name: sp_human:Q9UKR6
seq_documentation_block:
ID Q9UKR6 PRELIMINARY; PRT; 355 AA.
```



```

AC Q9UKR6;
DT 01-MAY-2000 (TrEMBLrel.13, Created)
DT 01-MAY-2000 (TrEMBLrel.13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel.19, Last annotation update)
DE KRUPPEL-LIKE FACTOR KLF.
DE Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC
PC TISSUE=LUNG.
RX MEDLINE=93989728; PubMed=10458913;
RA Wani M.A., Konkright M.D., Jeffries S., Hughes M.J., Lingrel J.B.;
RT "cDNA isolation, genomic structure, regulation, and chromosomal
RT localization of human lung kruppel-like factor.";
RL Genomics 60:78-86(1999).
DR EMBL: AF134053; AAD55891.1; -.
DR HSP: P08047; ISP2.
DR InterPro: IPR002965; P_rich_extensn.
DR InterPro: IPR000822; Znf-C2H2.
DR Pfam: PF00096; zf-C2H2; 3.
DR PRINTS: PR01217; PRICHEXTENS.
DR SMART: SM00355; Znf_C2H2; 3.
DR PROSITE: PS00026; ZINC_FINGER_C2H2_1; 3.
DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 3.
DR DNA-binding; Metal-binding; Zinc-finger.
KW
SQ SEQUENCE 355 AA; 37454 MW; CDD1735CC516DE24 CRC64;

alignment_scores:
    Quality: 129.00      Length: 223
    Ratio: 1.344        Gaps: 13
    Percent Similarity: 43.049      Percent Identity: 29.148

alignment_block:
US-09-462-480-2/rev x Q9UKR6

Align seg 1/1 to: Q9UKR6 from: 1 to: 355

523 CTGGACTACTTTCTCTCTTTA..... 503
|||||:|||||:|||||:
45 LeuAspPheIleLeuSerMetGlyLeuAspGlyLeuGlyAlaGluAla 61

502 .....CCTCTCCGCCAAATGTTGCCAGTCTTCGCGC 469
|||||:|||||:|||||:
61 aProGluProProProProProProProProAlaPheTyrTyrProGlu 78

468 CCGGGTGGCGC.....GGAAGTCGTGTGCA... 443
|||||:|||||:|||||:
78 roGlyAlaProProProTyrSerAlaProAlaGlyGlyLeuValSerGlu 94

442 ...TTACGGGAGCTACACGCTGCTCTTCGTCCTCCAGT...CGTCCTCG 399
|||||:|||||:|||||:|||||:
95 LeuLeuArgProGluLeuAspAlaProLeuGlyProAlaLeuHisGly 111

398 TCGTCTCTTCAGCTCTCGCGAGCGTCCGCGCGCACCA... 354
|||||:|||||:|||||:|||||:
111 gPheLeuAlaProProGlyArgLeuValLysAlaGluProProGlu 128

353 .....CCCGGGCTGGTGGAGCGCGC 333
|||||:|||||:|||||:
128 laAspGlyGlyGlyTyrGlyGlyCysAlaProGlyLeuThrArgGlyPro 144

332 .....GATTGGACCCCTGGCCCATCGCTCCC..... 306
|||||:|||||:|||||:
145 ArgGlyLeuLysArgGluGlyAlaProGlyProAlaAlaSerCysMet 161

305 ..GGACCCACCGGAGCGCGCCACCGCTCACCGAGTCCGGCACACGCG 257
|||||:|||||:|||||:
161 gGlyProGlyArgProProProProProProAspThrPro..... 174

256 CCGGCATCATCCGAGGGGGCAACCGGCTTTTCGATCAGCTGAGAC..... 213

```

```

37   GGCACCGGCGGGCAACCACCGCAGCAGAACCCGCCGAGATGGCGCT 86
    ||| |||||:::|||||::: |||
1146 GlyProGlyGlyAlaGlyProGlyGlyAlaGlyProGlyGlyAlaGlyPr 1162
    |||:::|||||::: |||
87   GCTGGCACCAAGTCGCTGTGCGAACCATCCGCTGGCTGGTGGATCAGGCC 136
    |||:::|||| | |||||:::||||
1162 oGlyGlyAlaGlyPro.....GlyGlyAlaGlyP 1172
    |||:::|||| | |||||:::||||
137  CCACGCGGGCGCGGGGCTGCTGCGCGCAGTGCTACTCGCGCAGGT 186
    |||:::|||| | |||:::
1172 roGlyGlyAlaGlyGlyAlaGlyGlySerGlyGlyAlaGly 1188
    |||:::|||| | |||:::
187  GGGTCTGTGACCCGACGCGCGCTCATGCTCAGCTG..... 222
    ||||| |||
1189 GlySer...GlyGlyThrThrIleileGluAspLeuAspIleThrIleAs 1204
    ||||| |||
223 ...ATCGAAAGCGGTGTGCCCTCG...GTGATCGCGCGGCTGTGTTG 265
    ::| |||||::: ||| |||||::: |
1204 pGlyAlaAspGlyProIleThrIleSerGluGluLeuProIleSerGlyA 1221
    ::| |||||::: ||| |||||::: |
266 CCGATGCTCGGTGACGGGTGGCGCGCTCCG.....GTGGTCCGGGA 309
    |||||:::|||||::: |||||:::|||||:::
1221 laGlyGlySerGlyProGlyGlyAlaGlyProGlyGlyValGlyProGly 1237
    |||||:::|||||::: |||||:::|||||:::
310 CGATGGCGCAGGTGTCGAATCCGGCGGCTCCACACGCCCCGGTCTGGT 359
    |||||:::|||||::: |||||:::|||||:::
1238 GlySerGlyProGlyGlyValGlyProGlyGlySerGlyProGlyGlyVa 1254
    |||||:::|||||::: |||||:::|||||:::
360 CGCGCGG 366
    |||||:::|||||:::
1254 lGlypro 1256

```

seq\_name: sp\_invertebrate:044358

```

seq_documentation_block:
ID      044358      PRELIMINARY;      PRT;      871 AA.
AC      044358;
DT      01-JUN-1998 (TremBLrel. 06, Created)
DT      01-JUN-1998 (TremBLrel. 06, Last sequence update)
DT      01-JUN-2001 (TremBLrel. 17, Last annotation update)
DT      01-JUN-2001 (TremBLrel. 17, Last annotation update)
DE      FLAGELLIFORM SILK PROTEIN (FRAGMENT).
GE      FLAG.
GN      GN.
OS      Nephila clavipes (Orb spider).
OC      OC.
OC      Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC      Araneomorphae; Entelegynae; Araneoldea; Tetragnathidae; Nephila.
OX      NCBTaxID=6915;
OX      [1]
RN      SEQUENCE FROM N.A.
RP      MEDLINE=98153262; PubMed=9480768;
RX      Hayashi C.Y., Lewis R.V.;
RA      "Evidence from flagelliform silk cDNA for the structural basis of
RT      elasticity and modular nature of spider silks.";
RL      J. Mol. Biol. 275:773-784(1998).
RL      EMBL; AF027972; AAC38846.1; -.
DR      InterPro; IPR000087; Collagen.
DR      InterPro; IPR000209; Peptidase_S8.
DR      PROSITE; PS00138; SUBTILASE_SER; UNKNOWN_1.
DR      NON_TER      871      871
SQ      SEQUENCE      871 AA; 71039 MW; 1FA1E3B7EC05983A CRG64;

```

```

alignment_scores:
  Quality: 128.00      Length: 132
  Ratio: 1.707         Gaps: 6
  Percent Similarity: 56.818      Percent Identity: 35.606

alignment_block:
  US-09-462-480-2 x 04358      ..

Align seg 1/1 to: 04358 from: 1 to: 871

```

37 GGCACCGGGCGGCGCACCCAGCCGACGAGGAAGCCGGCAGATGGGGCT 86

```
|||||
377 GlyProGlyGlyAlaGlyProGlyGlyAlaGlyProGlyGlyAlaGlyPr 393
87 GCTCGGCACACAGTCGCTGTCGACACCATCGCTGGTGGATCAGGCC 136
|||||
393 oGlyGlyTyGlyProGlyGlySerGlyPro.....GlyGlyAlaGlyP 408
|||||
137 CCACGGGGGGCGGGCGCTGCTGCGCGGAGTCGCTACCTGCGCGCAGGT 186
|||||
408 roSer....GlyAlaGlyLeuGlyAlaGly.....ProGlyGlyAla 421
|||||
187 GGGTCGTTGACCCGACCGCGCTGATGCTCAGCTGATCGAAAGCCGGT 236
|||||
422 Gly..... 422
237 TGCCCCCTCGGTGATCGCGGGCTGTTGCGGATCGTCGGTGACGGTG 286
|||||
423 .....LeuGlyGlyAlaGlyProGlyGlyAlaGlyThrSerG 435
287 GCGCGGCTCCG.....GTGGTCCGGAGCGATGGCGCGGTTCGCAA 330
|||||
435 lyAlaGlyProGlyGlyAlaGlyProGlyGlyAlaGlyGlnGlyAspAla 451
331 TCCGGCGGCTCCACAGGCC....GGTCTGTCGCGCGCGCACCGCTCG 376
|||||
452 GlyProGlyGlyAlaGlyArgGlyGlyAlaGlyArgGlyGlyValGlyAr 468
377 GCGAGGAGCGTGAAGACAGCAGCAGCAGTGGCAGGAGAGGA 422
468 gGlyGlyAlaGlyArgGlyGlyAlaGlyArgGlyGlyAlaArgGly 483
```

seq\_name: sp\_rodent:Q9WUE8

```
seq_documentation_block:
ID Q9WUE8 PRELIMINARY; PRT; 2087 AA.
AC Q9WUE8;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SHANK1A.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
[1]
SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY;
RX MEDLINE=99360650; PubMed=10433268;
RA Naisbitt S., Kim E., Tu J.C., Xiao B., Sala C., Valtschanoff J.,
RA Weinberg R.J., Worley P.F., Sheng M.;
RT "Shank, a novel family of postsynaptic density proteins that binds to
RT the NMDA receptor/PSD-95/GKAP complex and cortactin.";
RL Neuron 23:569-582(1999)
DR EMBL; AF131951; AAD29417.1; -.
DR HSSP; P00519; IABL.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR002965; P_rich_extensn.
DR InterPro; IPR001680; SAM.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00023; ank; 6.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00536; SAM; 1.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PR01217; PRICHEXTENS.
DR SMART; SM00248; ANK; 3.
DR SMART; SM00228; PDZ; 1.
DR SMART; SM00454; SAM; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50088; ANK_REPEAT; 3.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS50106; PDZ; 1.
DR PROSITE; PS50002; SH3; 1.
```

```
KW ANK repeat; Repeat.
SQ SEQUENCE 2087 AA; 218125 MW; 0955C33D7A7AA8F8 CRC64;

alignment_scores:
Quality: 127.50 Length: 124
Ratio: 1.903 Gaps: 9
Percent Similarity: 54.032 Percent Identity: 36.290

alignment_block:
US-09-462-480-2/rev x Q9WUE8 ..
Align seg 1/1 to: Q9WUE8 from: 1 to: 2087

353 CCGGG.....CTGGTGAGCGCGCGAT...TGCAGACCCCTG 319
|||||
1512 ProGlyProHisProLeuProAspProSerProAlaThrProLe 1528
318 GCCATCGCTCCCGACCC...ACCGAGCGCGCCACCCGTCACCGAGC 272
|||||
1528 uProAlaAlaProProAlaValAlaAlaProProThrLeuAspS 1545
271 ATCCGGCAACAGCC.....GCCGGCATCACC 246
:: |||:::
1545 erThrAlaSerSerLeuThrSerTyrAspSerGluValAlaThrLeuThr 1561
245 GAGGGGCAACCGCTTTTCGATCAGCTGAGACATCAGCGCGTGGGGT 196
:: |||:::
1562 GlnGlyAlaPro.....AlaAlaPr 1568
195 CAAGCACCCACCTGCCAGGT.....AGCGACTCCGCGCAGCAGGC 152
:: |||:::
1568 oGlyAspProAlaProGlyProProAlaProAlaProAlaProp 1585
151 CCGCGCGCGCTGGGCTGATCCACCGCAGCGGATGTTCCACAGC 102
|||||
1585 roAlaProGlnProGlyProAspProProGly.....ThrAspSer 1599
101 GGACTG.....GTCCCGCAGCAGG.....CCCATCTGCGC 73
|||||
1600 GlylleGluGluValAspSerSerSerSerSerHisProLeuGluTh 1616
72 GGCTTCCTCGTCGCTGGTGGTGG 51
1616 rIleSerSerAlaSerThrLeu 1623
```

seq\_name: sp\_rodent:Q9WU13

```
seq_documentation_block:
ID Q9WU13 PRELIMINARY; PRT; 2158 AA.
AC Q9WU13;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SYNAPTIC SAPAP-INTERACTING PROTEIN SYNAMON.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
[1]
SEQUENCE FROM N.A.
RX MEDLINE=99419021; PubMed=10488079;
RA Yao I., Hata Y., Hirao K., Deguchi M., Ide N., Takeuchi M., Takai Y.;
RT "Synamon, a novel neuronal protein interacting with synapse-associated
RT protein 90/postsynaptic density-95-associated protein.";
RL J. Biol. Chem. 274:27463-27466(1999).
DR EMBL; AF102855; AAD04569.2; -.
DR HSSP; P00519; IABL.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR002965; P_rich_extensn.
DR InterPro; IPR001680; SAM.
DR InterPro; IPR001452; SH3.
```

DR Pfam; PF00023; ank; 6.  
DR Pfam; PF00595; PDZ; 1.  
DR Pfam; PF00536; SAM; 1.  
DR Pfam; PF00018; SH3; 1.  
DR PRINTS; PR01217; PRICHTEXTENS.  
DR SMART; SM00248; ANK; 3.  
DR SMART; SM00228; PDZ; 1.  
DR SMART; SM00454; SAM; 1.  
DR SMART; SM00326; SH3; 1.  
DR PROSITE; PS50088; ANK\_REPEAT; 3.  
DR PROSITE; PS50297; ANK\_REPEAT\_REGION; 1.  
DR PROSITE; PS50106; PDZ; 1.  
DR PROSITE; PS50002; SH3; 1.  
KW ANK repeat; Repeat.  
SQ SEQUENCE 2158 AA; 225519 MW; 2AB6D53B5F1A4407 CRC64;

alignment\_scores:      Quality: 127.50      Length: 124  
                         Ratio: 1.903                      Gaps: 9  
Percent Similarity: 54.032      Percent Identity: 36.290

alignment\_block:

US-09-462-480-2/rev x Q9WU13      ..

Align seg 1/1 to: Q9WU13 from: 1 to: 2158

```
353 CCCGGG.....CTGGTGGAGCCGCCGAT...TCCGAACCCGTG 319
||||| ||| :|||||: : : : |||
1593 ProGlyProHisProLeuProAspProSerProAlaThrProLe 1599
||| |||| ||| :|||||: : : : |||
318 GCCATCGCTCCCGACCC...ACCGAGCGGCCGCCCGTCACCGACG 272
||| |||| ||| :|||||: : : : |||
1599 uProAlaAlaProProAlaValAlaAlaProProThrLeuAsps 1616
271 ATCCGGCAACAGCC.....GCCGGCATCACC 246
:: |||: : : : ||| :|||
1616 erThrAlaSerSerLeuThrSerTyrAspSerGluValAlaThrLeuThr 1632
::||| |||
245 GAGGGGCAACCGGCTTTTCGATCAGCTGACATCAGCGCGTGCGGGT 196
::||| |||
1633 GlnGlyAlaPro.....AlaAlaPr 1639
195 CAACGACCCACTCGCCAGGT.....AGCGACTCCGCGCGCAGCAGC 152
::||| ||||| ||||| : : : |||
1639 oGlyAspProAlaProGlyProAlaProAlaAlaProAlaProp 1656
151 CCGCGCGCGCGCTGGGCGCTGATCCACCGCCGCGGATGTTGACAGC 102
||||| ||||| ||||| : : : |||
1656 roAlaProGlnProGlyProAspProProGly.....ThrAspSer 1670
101 GGACTG.....GTCCCGAGCAGG.....CCCATCTGCGC 73
|||: : ||| |||||
1671 GlyIleGluGluValAspSerSerArgSerSerSerAspHisProLeuGluTh 1687
72 GGCTTCCTCGTCGGCTGGTTG 51
: |||||: : |||
1687 rIleSerAlaSerThrLeu 1694
```

OM of: US-09-462-480-1 to: A\_Geneseq\_032802.\* out\_format : pfs  
Date: Jul 22, 2002 1:22 AM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

## Command line parameters:

-MODEL=framet.n2p.model -DEV=xlh  
-O=/cn2.1/USPTO.spool/US09462480/runat\_18072002.164418.19413/app\_query.fasta\_1.2850  
-DB=A\_Geneseq\_032802\_QFMT-fastan -SUFFIX=rag -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-CGAPOP=4.500 -CGAPEXT=0.050 -YGAPOP=10.000 -YGAPEXT=0.500  
-FCGAPOP=6.000 -FCGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blotsum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09462480\_cgn1\_1\_57 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

## Search information block:

Query: US-09-462-480-1  
Query length: 1277  
Database: A\_Geneseq\_032802.\*  
Database sequences: 747574  
Database length: 111073796  
Search time (sec): 134.850000

## score\_list:

Sequence	Strd Orig	ZScore	Escore	Len	Documentation
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:AAW32452 +	713.00	1027.91	1.1e-49		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:AAW32384 +	713.00	1027.91	1.1e-49		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW81705 +	713.00	1027.91	1.1e-49		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW64338 +	713.00	1027.91	1.1e-49		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW39135 +	713.00	1027.91	1.1e-49		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW38992 +	713.00	1027.91	1.1e-49		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:AAW39844 +	713.00	1027.91	1.1e-49		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW2929 +	713.00	1027.83	1.1e-49		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW21946 +	713.00	1027.83	1.1e-49		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW29888 +	512.50	741.37	1.9e-33		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW30705 +	492.00	718.28	7.3e-32		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW22942 +	487.50	697.48	2.6e-31		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW21962 +	487.50	697.48	2.6e-31		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW81746 +	487.00	689.76	3.5e-31		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW64379 +	487.00	689.76	3.5e-31		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW32063 +	487.00	689.76	3.5e-31		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW39224 +	487.00	689.76	3.5e-31		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW39081 +	487.00	689.76	3.5e-31		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW39033 +	486.00	689.76	3.5e-31		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW81706 +	486.00	709.51	2.2e-31		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW64339 +	486.00	709.51	2.2e-31		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW39136 +	486.00	709.51	2.2e-31		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW38993 +	486.00	709.51	2.2e-31		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:AAW35218 +	486.00	709.51	2.2e-31		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:AAW39845 +	486.00	709.51	2.2e-31		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:AAW21901 +	485.00	684.76	5.5e-31		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW22943 +	482.00	689.47	7.3e-31		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW321963 +	482.00	689.47	7.3e-31		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:AAW11494 +	479.00	699.81	8.2e-31		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW29890 +	479.00	699.81	8.2e-31		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW29788 +	479.00	699.81	8.2e-31		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:AAW35219 +	479.00	699.81	8.2e-31		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:AAW32444 +	462.00	674.97	2.0e-29		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:AAW32376 +	462.00	674.97	2.0e-29		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW81747 +	462.00	674.97	2.0e-29		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW64321 +	462.00	674.97	2.0e-29		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW32097 +	462.00	674.97	2.0e-29		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW39118 +	462.00	674.97	2.0e-29		
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW38981 +	462.00	674.97	2.0e-29		

/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:AAW32454 + 392.00 574.44 9.4  
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:AAW32386 + 392.00 574.44 9.4  
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW81707 + 392.00 574.44 9.4  
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW64340 + 392.00 574.44 9.4  
/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAW39137 + 392.00 574.44 9.4  
seq\_name: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:AAW32452

## seq\_documentation\_block:

ID AAW32452 standard; Protein; 368 AA.  
XX  
AC AAW32452;  
XX  
DT 09-JAN-1998 (first entry)  
XX  
DE Mycobacterium tuberculosis antigen Tb37-FL.  
XX  
KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;  
KW skin testing; M.tuberculosis.  
XX  
OS Mycobacterium tuberculosis.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 10 /note= "Any amino acid"  
XX  
PN WO9709428-A2.  
XX  
PD 13-MAR-1997.  
XX  
PF 30-AUG-1996; 96WO-US14674.  
XX  
PR 12-JUL-1996; 96US-0680574.  
PR 01-SEP-1995; 95US-0523436.  
PR 22-SEP-1995; 95US-0533634.  
PR 22-MAR-1996; 96US-0620874.  
PR 05-JUN-1996; 96US-0659683.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Campos-neto A, Dillon DC, Houghton R, Reed SG, Skelky YAW;  
PI Twardzik DR, Vedvick TH;  
XX  
DR WPI; 1997-192903/17.  
XX  
PT New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are  
PT useful in vaccines for prevention or treatment of tuberculosis, also  
PT for diagnosis  
XX  
PS Example 3; Page 146-147; 168pp; English.  
XX  
CC A new immunogenic polypeptide has been developed comprising an  
CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or  
CC its variant differing only in conservative substitutions and/or  
CC modifications). The present sequence represents a M.tuberculosis  
CC antigen, Tb37-FL The immunogenic protein, and fusion proteins  
CC containing one or more of the proteins or one of the proteins plus  
CC ESAT-6, are useful in vaccines, preferably when formulated with a  
CC non-specific adjuvant, to induce an immune response against  
CC M.tuberculosis (for treatment or prevention).  
XX  
SQ Sequence 368 AA;

alignment\_scores:  
Quality: 713.00 Length: 143  
Ratio: 5.057 Gaps: 0

Percent Similarity: 98.601 Percent Identity: 96.503

## alignment\_block:

US-09-462-480-1 x AAW32452 ..

Align seg 1/1 to: AAW32452 from: 1 to: 368

```
1 CTGACGAGGTGACGTCGTGTTACCCAGGTGGCGGCACCGCGCGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGlyG 242
|||||
51 CAACCCAGCGGAGGAGGAGCGGAGATGGCGCTGTCGGCAGCAGTC 100
|||||
242 yAsnProAlaAspGluGluAlaGlnMetGlyLeuLeuGlyThrSerP 259
|||||
101 CGCTGTGCAACCATCCGCTGGTGGTATCAGGCCAGCCAGCGCGCGG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275
|||||
151 GCGTCTGCTGCGCGGAGTGGTCTGCTGCTGCTGCTGCTGCTGCTGCT 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292
|||||
201 CACGCGCTGATCTCTACGTATCGAAGAGCGGTTGCCCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuIleGlnLysProValAlaProSerValM 309
|||||
251 TGCCGCGCGCTGTCGCGATCGCTGGTACGCGGTGGCGCGCTCGGTG 300
|||||
309 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
|||||
301 GTCGCGGAGGATGGCGGAGGTTCGCAATCCGCGCGCTCCACGAGCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
|||||
351 GCGTCTGTCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAspG 359
|||||
401 AGGAGACTGGGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 429
|||||
359 luAspAspTrpAspGluGluAspAspTrp 368
|||||
```

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:AAW32384

seq\_documentation\_block:

ID\_AAW32384 standard; Protein; 368 AA.

AC\_AAW32384;

13-JAN-1998 (first entry)

Mycobacterium tuberculosis antigen Tb37-FL.

Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;  
skin testing; M.tuberculosis.

Mycobacterium tuberculosis.

Key Location/Qualifiers

Misc\_difference 10 /note= "Any amino acid"

WO9709429-A2.

13-MAR-1997.

30-AUG-1996; 96WO-US14675.

12-JUL-1996; 96US-0680573.

01-SEP-1995; 95US-0523435.

22-SEP-1995; 95US-0532136.

22-MAR-1996; 96US-0620280.

03-JUN-1996; 96US-0658800.

(CORI-) CORIXA CORP.

Campes-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;

PI\_Twardzik DR, Vedvick TH;

```
XX WPI; 1997-192904/17.
XX DR
XX PT New immunogenic polypeptide(s) from soluble M. tuberculosis antigens
XX - useful for diagnosis of M. tuberculosis infection
XX PS
XX PS Example 3; Page 159-161; 190pp; English.
XX CC A new immunogenic polypeptide has been developed comprising an
XX immunogenic part of a soluble Mycobacterium tuberculosis antigen (or
XX its variant differing only in conservative substitutions and/or
XX modifications). The present sequence represents a M.tuberculosis
XX antigen, Tb37-FL. The immunogenic polypeptide can be used to diagnose
XX M.tuberculosis infection by forming complexes with specific
XX antibodies in the sample. Fragments of DNA encoding the immunogenic
XX polypeptide can be used as diagnostic primers or probes and agents
XX that bind to the antigen, especially monoclonal antibodies or
XX equivalent polyclonal antibodies, are also used for diagnosis.
XX SQ Sequence 368 AA;
```

alignment\_scores:

Quality: 713.00 Length: 143

Ratio: 5.057 Gaps: 0

Percent Similarity: 98.601 Percent Identity: 96.503

alignment\_block:

US-09-462-480-1 x AAW32384

Align seg 1/1 to: AAW32384 from: 1 to: 368

```
1 CTGACGAGGTGACGTCGTGTTACCCAGGTGGCGGCACCGCGCGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGlyG 242
|||||
51 CAACCCAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 100
|||||
242 yAsnProAlaAspGluGluAlaGlnMetGlyLeuLeuGlyThrSerP 259
|||||
101 CGCTGTGCAACCATCCGCTGGTGGTATCAGGCCAGCCAGCGCGCGG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275
|||||
151 GCGCTGTCTGCGCGGAGTGGCTACCTGGCGCAGGTGGGTCTGTTGACCCG 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292
|||||
201 CACGCGCTGATCTCTACGTATCGAAGAGCGGTTGCCCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuIleGlnLysProValAlaProSerValM 309
|||||
251 TGCCGCGCGCTGTCGCGATCGCTGGTACGCGGTGGCGCGCTCGGTG 300
|||||
309 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
|||||
301 GGTCTGTCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
|||||
351 GCGTCTGTCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAspG 359
|||||
401 AGGAGACTGGGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 429
|||||
359 luAspAspTrpAspGluGluAspAspTrp 368
|||||
```

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW81705

seq\_documentation\_block:

ID\_AAW81705 standard; Protein; 368 AA.

XX

AC AAW81705;  
 XX 27-JAN-1999 (first entry)  
 XX  
 DE M. tuberculosis immunogenic polypeptide Tb37-FL.  
 XX  
 XX Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;  
 KW vaccine; pharmaceutical; infection; diagnosis.  
 XX  
 OS Mycobacterium tuberculosis.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 10  
 FT /label= unknown  
 XX  
 XX WO9816646-A2.  
 XX  
 PD 23-APR-1998.  
 XX  
 XX 07-OCT-1997; 97WO-US18293.  
 XX  
 PR 13-MAR-1997; 97US-0818112.  
 PR 11-OCT-1996; 96US-0730510.  
 XX  
 XX (CORI-) CORIXA CORP.  
 PA  
 XX Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;  
 PI Reed SG, Skeiky YAW, Twardzik DR, Vedwick TS;  
 XX  
 XX WPI; 1998-261042/23.  
 DR  
 XX Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used  
 PT to develop products for the detection of M. tuberculosis infection  
 PT and for diagnosis, treatment and prevention of tuberculosis  
 XX  
 PS Example 3B; Page 137-138; 230pp; English.  
 XX  
 CC This sequence represents an immunogenic portion of a soluble  
 CC Mycobacterium tuberculosis (MT) antigen which can be used in a method  
 CC for inducing protective immunity against tuberculosis (TB). This  
 CC sequence can be formulated into vaccines and/or pharmaceutical  
 CC compositions for immunising against M. tuberculosis infection or may  
 CC be used for the diagnosis of tuberculosis.  
 XX  
 SQ Sequence 368 AA;  
 alignment\_scores:  
 Quality: 713.00 Length: 143  
 Ratio: 5.057 Gaps: 0  
 Percent Similarity: 98.601 Percent Identity: 96.503  
 alignment\_block:  
 US-09-462-480-1 x AAW81705 ..  
 Align seq 1/1 to: AAW81705 from: 1 to: 368  
 1 CTGCAGCAGGTGAGCTGTTGTCAGCCAGGTGGCGGCACCGCGCGG 50  
 |||||||||||||||||||||||||||||||||||||||||||||||  
 226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGly 242  
 51 CAACCCAGCCGACGAGGAGCGCGCAGATGGCGCTCGGCACCAAGTC 100  
 |||||||||||||||||||||||||||||||||||||||||||||||  
 242 YASNProAlaAspGluGlnAlaGlnMetGlyLeuLeuGlyThrSerP 259  
 101 CGCTGTGACACCATCCGCTGGTGGATGATGAGCCCGCGCGCGCG 150  
 |||||||||||||||||||||||||||||||||||||||||||||||  
 259 roLeuSerAsnHisProLeuAlaGlySerGlyProSerAlaGlyAla 275  
 151 GGCTGCTGCGCGGAGTGGCTACCTGGCGCAGGTGGTGGTGGTGGC 200  
 |||||||||||||||||||||||||||||||||||||||||||||||  
 276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlySerLeuThrAr 292

201 CACGCCGCTGATGCTCAGCTGATCGAAAAAGCGGTTGCCCTCGGTGA 250  
 |||||||||||||||||||||||||||||||||||||||||||||||  
 292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309  
 251 TGCCGGCGGCTGTTGCCGGATCGTTCGGTGACGGTGCGCGCTCGCGTG 300  
 |||||||||||||||||||||||||||||||||||||||||||||||  
 309 etProAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325  
 301 GGTCCGGGAGCGATGGCCAGGTTTCGAATCGCGGGCTCCACCGCCC 350  
 ||| ||||||||||||||||||||||||||||||||||||||||  
 326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342  
 351 GGTCTGTGTCGCCGCCGACCGCTCGCGCAGGAGCGGTGAAGAAGACGAC 400  
 |||||||||||||||||||||||||||||||||||||||||||||||  
 342 OGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAspG 359  
 401 AGGACGACTGGGAGCAGAGGAGGACTGG 429  
 ||||||||||||||||||||||||||||||||||||||||  
 359 luAspAspTrpAspGluGluAspAspTrp 368  
 seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW64338  
 seq\_documentation\_block:  
 ID AAW64338 standard; Protein: 368 AA.  
 XX  
 AC AAW64338;  
 XX  
 DT 09-NOV-1998 (first entry)  
 XX  
 DE Mycobacterium tuberculosis antigen Tb37-FL.  
 XX  
 KW Tuberculosis; infection; diagnosis; antigen; TB37-FL.  
 XX  
 OS Mycobacterium tuberculosis strain H37Rv.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 10  
 FT /note= "unidentified"  
 XX  
 XX WO9816645-A2.  
 PN  
 XX 23-APR-1998.  
 PD  
 XX 07-OCT-1997; 97WO-US18214.  
 XX  
 XX 13-MAR-1997; 97US-0818111.  
 PR  
 PR 11-OCT-1996; 96US-0729622.  
 XX  
 XX (CORI-) CORIXA CORP.  
 PA  
 XX Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;  
 PI Reed SG, Skeiky YAW, Twardzik DR, Vedwick TS;  
 XX  
 XX WPI; 1998-251292/22.  
 DR  
 XX  
 XX New isolated Mycobacterium tuberculosis polypeptides and DNA - used  
 PT to develop products for the detection of M. tuberculosis infection  
 PT and diagnosis of tuberculosis  
 XX  
 PS Example 3; Page 143-144; 250pp; English.  
 XX  
 CC This polypeptide comprises Mycobacterium tuberculosis antigen  
 CC Tb37-FL. It is encoded by genomic DNA isolated from a M.  
 CC tuberculosis strain H37Rv genomic library using a probe from  
 CC clone Tb38-1 (see AAV44384). The invention relates to compositions  
 CC and methods for diagnosing tuberculosis. It provides polypeptides  
 CC (see AAW64291-W64379) comprising an antigenic portion of a soluble  
 CC M. tuberculosis antigen, or an immunogenic portion of an M.  
 CC tuberculosis antigen, as well as DNA sequences encoding such  
 CC polypeptides, recombinant expression vectors and transformed or  
 CC transfected host cells. Also claimed are methods and diagnostic  
 CC kits for detecting M. tuberculosis infection in a patient using  
 CC these polypeptides, antibodies or oligonucleotide probes and





401 AGGACGACTGGAGCAAGAGCAGCACTGG 429  
359 luAspAspTTPAspGluGluAspAspTrp 368

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AA1999

seq\_documentation\_block:

ID AAY38992 standard; Protein; 368 AA.

XX

AC AAY38992;

XX

DT 05-NOV-1999 (first entry)

XX

DE M. tuberculosis recombinant antigen protein Tb37-FL.

XX

KW Antigen; diagnosis; detection; infection; antibody; immunisation;

KW vaccine; immunity.

XX

OS Mycobacterium tuberculosis.

XX

PN WO9942118-A2.

XX

PD 26-AUG-1999.

XX

PF 17-FEB-1999; 99WO-US03265.

XX

PR 05-MAY-1998; 98US-0072596.

XX

PR 18-FEB-1998; 98US-0024753.

XX

PA (CORI-) CORIXA CORP.

XX

PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;

PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;

DR WPI; 1999-527416/44.

XX

PT New polypeptide comprising antigenic portions of M. tuberculosis

XX

PS Example 3; Page 177-179; 323pp; English.

XX

CC This invention describes novel recombinant antigens and their encoding

CC nucleic acids derived from Mycobacterium tuberculosis. The novel

CC polypeptides are useful for detecting M. tuberculosis infection in a

CC biological sample by detecting antibodies which bind with the

CC polypeptides, and are useful as vaccines for immunizing against

CC M. tuberculosis infection. The new detection methods are needed as

CC current vaccination strategies do not provide 100% immunity.

XX

SQ Sequence 368 AA;

alignment\_scores:

Quality: 713.00

Length: 143

Ratio: 5.057

Gaps: 0

Percent Similarity: 98.601

Percent Identity: 96.503

alignment\_block:

US-09-462-480-1 x AAY38992

..

Align seg 1/1 to: AAY38992

from: 1 to: 368

1 CTGCAGCAGTGTACGCTGTTGTTTCAGCCAGGTGGCGGCACCGCGCGG 50

|||||

226 leuGlnGlnValThrSerPheSerGlnValGlyThrGlyGlyG1 242

|||||

51 CAACCCAGCCGACGAGAGCCGCGCAGATGGGCTGCTCGGCACCAAGTC 100

|||||

242 YAsnProAlaaspGluGluAlaGlnMetGlyLeuLeuGlyThrSerP 259

|||||

101 GCTGTGCAACATCCGCTGCTGTGATCAGGCCCGCCAGCGCGGCGCG 150

|||||

259 rOleuSerAsnHISProLeuAlaGlySerGlyProSerAlaGlyAla 275

151 GGCTCTGCTGCGCGAGTGGCTACCTGGCGGAGTGGTGGTGGTGGCCG 200  
|||||  
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292  
201 CAGCCGCTGATGCTCAGCTGATCGAAAGCCGGTGGCCCTCGGTGA 250  
|||||  
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309  
251 TGGCGGCGCTGTTGGCGGATCGCTGACGGGTGGCGGCTGCCGCTG 300  
|||||  
309 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325  
301 GGTCCGGGAGCGATGGCCAGGTTCCAAATCCGGCGGCTCCACCGCC 350  
|||||  
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342  
351 GGTCTGCTGCGCGCGCGCGCTCGCGCAGGAGCGGTGAAGAAGACGACG 400  
|||||  
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAspG 359  
401 AGGACGACTGGAGCAAGAGCAGCACTGG 429  
|||||  
359 luAspAspTTPAspGluGluAspAspTrp 368

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:AA19844

seq\_documentation\_block:

ID AAB19844 standard; Protein; 368 AA.

XX

AC AAB19844;

XX

DT 05-MAR-2001 (first entry)

XX

DE Mycobacterium tuberculosis protein MTBN3.

XX

KW MTBN3; tuberculosis; BCG; vaccine; infection; diagnosis.

XX

OS Mycobacterium tuberculosis.

XX

PN WO2000066157-A1.

XX

PD 09-NOV-2000.

XX

PF 04-MAY-2000; 2000WO-US12257.

XX

PR 04-MAY-1999; 99US-0132505.

XX

PA (PUBL-) PUBLIC HEALTH RES INST NEW YORK.

XX

PI Gennaro ML;

XX

DR WPI; 2001-007153/01.

XX

DR N-PSDB; AAA89037.

XX

PT Novel polypeptide encoded by open reading frames present in

PT Mycobacterium tuberculosis genome and not by the BCG strain of M.

PT bovis, useful as vaccine and for diagnosing tuberculosis infection

XX

PS Claim 11; Fig 1; 35pp; English.

XX

CC The present sequence is that of the Mycobacterium tuberculosis

CC MTBN3 protein. This is 1 of 8 proteins, i.e. MTBN1-8 (see

CC AAB19842-49), encoded by 8 open reading frames (see AAA89035-42)

CC identified as being present in the genome of M. tuberculosis but

CC absent from the genome of the BCG strain of Mycobacterium bovis.

CC MTBN1-8 represent reagents that are useful in discriminating between

CC M. tuberculosis and BCG and, in particular, for diagnostic methods

CC which discriminate between exposure of a subject to M. tuberculosis

CC and vaccination with BCG. The invention features these MTBN

CC polypeptides, functional fragments of them, DNA encoding them,

CC vectors, transformed cells, and diagnostic, therapeutic, and

CC prophylactic (vaccine) methods, including genetic vaccination

CC methods.

```
XX SQ Sequence 368 AA;
alignment_scores:
  Quality: 713.00 Length: 143
  Ratio: 5.057 Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503
alignment_block:
US-09-462-480-1 x AAB19844 ..
Align seg 1/1 to: AAB19844 from: 1 to: 368
1 CTGCAGCAGGTGACGTCTGTTTCAGCCAGGTGGCGGCACCGCGCGG 50
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGlyG 242
51 CAACCCAGCCGACGAGGAGCCGCGCAGATGGCTGCTCGGCACCATC 100
242 YAsnProAlaAspGluGluAlaGlnMetGlyLeuLeuGlyThrSerP 259
101 CGCTGTCAACCATCCGCTGCTGCTGATCAGCCCGCCAGCGCGGCG 150
259 roLeuSerAsnHisProLeuAlaGlySerGlyProSerAlaGlyAla 275
151 GGCCTGTCTGCGCGGAGTCTGCTACCTGGCGCAGGTGGTGGTTC 200
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThr 292
201 CACGCGCTGATGCTCTAGCTGATCGAAGCCGGTTCGCCCTCGGTGA 250
292 gThrProLeuMetSerGlnLeuLeuGluLysProValAlaProSerVal 309
251 TGCCGGCGGCTGTTGCGGATCGTCTGCTGAGGTGGCGCGCTCCG 300
309 etProAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
301 GGTCCGGAGCGATGGCCAGGTTTCGAATCCGCGCGCTCCACAGCC 350
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrAr 342
351 GGGTCTGGTCCGCGCGCAGCGCTCGCGCAGGAGCGGTGAAGAAGC 400
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAsp 359
401 AGGACGACTGGGACGAGAGGAGGACTGG 429
359 LuAspAspTrrAspGluGluAspTrrp 368
seq_name: /STDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:AAW72929
seq_documentation_block:
ID AAW72929 standard; Protein; 371 AA.
XX AC AAW72929;
XX DT 21-JAN-1999 (first entry)
XX DE Mycobacterium tuberculosis antigen RDI-ORF5.
XX KW Mycobacterium tuberculosis; antigen; vaccine; immunological;
XX KW immunogen; infection.
XX OS Mycobacterium tuberculosis.
XX PN WO9841119-Al.
XX PD 08-OCT-1998.
XX PF 01-APR-1998; 98WO-DK00132.
XX PF 05-JAN-1998; 98US-0070488.
XX PR
```

```
PR 02-APR-1997; 97DK-0000376.
PR 18-APR-1997; 97US-0044624.
PR 10-NOV-1997; 97DK-0001277.
XX (STAT-) STATENS SERUM INST.
XX Andersen P, Florio W, Nielsen R, Oettinger T, Rasmussen PB;
PI Rosenkrands I, Weidlingh K;
XX WPI: 1998-542705/46.
DR N-PSDB; AAV63939.
XX
PT New isolated mycobacteria polypeptides and nucleic acids - used for
PT developing products for the diagnosis of or vaccination against
PT mycobacterial infections, particularly tuberculosis
XX Claim 1; Page 200-202; 163pp; English.
XX
CC The present sequence represents a Mycobacterium tuberculosis protein.
CC Products from the present invention, which describes protein fragments
CC and nucleic acid fragments derived from M.tuberculosis, can be used in
CC the detection of and prevention of mycobacterial infections. In
CC particular, the proteins and nucleic acids can be used for the diagnosis
CC of or vaccination against tuberculosis caused by M. tuberculosis,
CC M. africanum or M. bovis.
XX
SQ Sequence 371 AA;
alignment_scores:
  Quality: 713.00 Length: 143
  Ratio: 5.057 Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503
alignment_block:
US-09-462-480-1 x AAW72929 ..
Align seg 1/1 to: AAW72929 from: 1 to: 371
1 CTGCAGCAGGTGACGTCTGTTTCAGCCAGGTGGCGGCACCGCGCGG 50
229 LeuGlnGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGlyG 245
51 CAACCCAGCCGACGAGGAGCCGCGCAGATGGCTGCTCGGCACCATC 100
245 YAsnProAlaAspGluGluAlaGlnMetGlyLeuLeuGlyThrSerP 262
101 CGCTGTCAACCATCCGCTGCTGCTGATCAGCCCGCCAGCGCGGCG 150
262 roLeuSerAsnHisProLeuAlaGlySerGlyProSerAlaGlyAla 278
151 GGCCTGTCTGCGCGGAGTCTGCTACCTGGCGCAGGTGGTGGTTC 200
279 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThr 295
201 CACGCGCTGATGCTCTAGCTGATCGAAGCCGGTTCGCCCTCGGTGA 250
295 gThrProLeuMetSerGlnLeuLeuGluLysProValAlaProSerVal 312
251 TGCCGGCGGCTGTTGCGGATCGTCTGCTGAGGTGGCGCGCTCCG 300
312 etProAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 328
301 GGTCCGGAGCGATGGCCAGGTTTCGAATCCGCGCGCTCCACAGCC 350
329 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrAr 345
351 GGGTCTGGTCCGCGCGCAGCGCTCGCGCAGGAGCGGTGAAGAAGC 400
345 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAsp 362
401 AGGACGACTGGGACGAGAGGAGGACTGG 429
```

362 luAspAspTrpAspGluAspAspTrp 371

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAV21946

seq\_documentation\_block:

ID AAY21946 standard; Protein; 371 AA.

AC AAY21946;

DT 06-SEP-1999 (first entry)

DE Amino acid sequence of antigen RD1-ORF5.

XX Immunogenic; Mycobacterium tuberculosis; immune response; infection;  
KW tuberculosis; fusion polypeptide; T-cell epitope; ESAT-6; MPT59; TB;  
KW pharmaceutical; vaccination; M. africanus; M. bovis; CFP7A; CFP30A;  
KW CFP7B; CFP19; CFP27; CFP30A; RD1-ORF; CFP10A; CFP16; CFP19; CFP23;  
KW CFP25A; CFP30B; CFP7B.

XX Mycobacterium tuberculosis.

OS WO9924577-A1.

PN 20-MAY-1999.

PD 08-OCT-1998; 98WO-DK00438.

PF 01-APR-1998; 98WO-DK00132.

PR 10-NOV-1997; 97DK-0001277.

PR 05-JAN-1998; 98US-0070488.

XX (STAT-) STATENS SERUM INST.

XX Andersen P, Skjot R;

XX WPI; 1999-347282/29.

XX N-PSDB; AAX81046.

XX New immunogenic fragment of Mycobacterium tuberculosis

XX Example 2; Page 219-220; 265pp; English.

XX The invention describes a substantially pure immunogenic polypeptide  
CC fragment (1) from Mycobacterium tuberculosis that is able to evoke a  
CC protective immune response against infections by mycobacteria belonging  
CC to the tuberculosis complex. The invention provides a (1) fusion  
CC polypeptide comprising at least one polypeptide fragment (1) and at least  
CC one fusion partner; (2) a fusion polypeptide fragment comprising a T-cell  
CC epitope from M. tuberculosis protein ESAT-6, or MPT59 and a second  
CC different amino acid sequence from M. tuberculosis, and/or including a  
CC sequence which protects the first amino acid sequence from in vivo  
CC degradation or post-translational processing; (3) a nucleic acid fragment  
CC that encodes the above polypeptides. The polypeptides and nucleic acid  
CC are useful as pharmaceuticals, for diagnosis of and as antigens for  
CC vaccination against TB caused by Mycobacterium tuberculosis, africanus or  
CC bovis. The polypeptides are also useful for diagnosing ongoing or  
CC previous sensitization in an animal with bacteria belonging to the  
CC tuberculosis complex. The invention also describes the use of CFP7A or  
CC CFP30A or a T-cell epitope of for the induction of a strong immune  
CC response in a mammal; use of CFP7B, CFP19 or MPT59-ESAT6 or a T-cell  
CC epitope of for diagnosis of TB in a mammal by performing a DTH type skin  
CC test; use of CFP27, CFP30A, RD1-ORF2, RD1-ORF3, RD1-ORF5, MPT59-ESAT6,  
CC ESAT6-MPT59, CFP10A, CFP16, CFP19, CFP23, CFP25A, CFP30B, CFP7B or a T-  
CC cell epitope of for the preparation of an immunological composition; and  
CC for the preparation of a subunit vaccine.

XX Sequence 371 AA;

alignment\_scores:

Quality: 713.00 Length: 143  
Ratio: 5.057 Gaps: 0  
Percent Similarity: 98.601 Percent Identity: 96.503

alignment\_block:

US-09-462-480-1 x AAY21946 ..

Align seg 1/1 to: AAY21946 from: 1 to: 371

```

1 CTGCAGCAGGTGACGTCTGTTTCAGCCAGGTGGCGGCACCCGCGCGG 50
|||||
229 LeuGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGlyG 245
51 CAACCCAGCCGACGAGGAGGAGCCGCGCAGATGGGCTGCTGGCACCAGTC 100
|||||
245 YasnProAlaAspGluGluAlaAlaGlnMetGlyLeuLeuGlyThrSerP 262
101 CGCTGTCAACCATCCGCTGGTGGTATCAGCCAGCCAGCCGCGCGCG 150
|||||
262 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 278
151 GGCCTGTGCGCGCGGAGTCTACCTGGCGCAGGTGGTGGTCTGTGACCCG 200
|||||
279 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 295
201 CACGCCGCTGATGTCCTCAGCTGATCGAAAGCCGTTGCCCCCTCGGTGA 250
|||||
295 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 312
251 TGCCGCGCGCTGTTTGGCGGATCGTGGTACGGTGGCGCGCTCCCGGTG 300
|||||
312 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 328
301 GGTCGCGGAGCATGGCCAGGTTCGCAATCCGCGCGCTCCACCCAGCCC 350
|||||
329 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 345
351 GGGTCTGTCGCGCGCGCGCGCTCGCGCAGGCGGTGAAGAACAGCAGCG 400
|||||
345 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspSpG 362
401 AGGACGACTGGGACGAAAGAGGACGACTGG 429
362 luAspAspTrpAspGluAspAspTrp 371

```

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:AAV29888

seq\_documentation\_block:

ID AAY29888 standard; Protein; 196 AA.

XX AAY29888;

XX 18-NOV-1999 (first entry)

XX Mycobacterium tuberculosis ESAT-6 protein sequence.

XX Mycobacterial; lactic acid bacterium; diagnosis; skin test; vaccine;  
KW delayed type hypersensitivity; DTH; ESAT-6 homodimer; tuberculosis;  
KW interferon-gamma release.

XX Mycobacterium tuberculosis.

XX WO9945119-A2.

XX 10-SEP-1999.

XX 05-MAR-1999; 99WO-DK00109.

XX 06-MAR-1998; 98DK-0000306.

XX 06-MAR-1998; 98US-0077105.

XX (STAT-) STATENS SERUM INST.

XX Jensen CL, Folkersen J;

XX WPI; 1999-551043/46.





211 gAsnAspProThrGlnGlnMileProLysLeuValAlaAsnAsnThrArgL 228  
569 AGGTAATTCGAGCGGATCT ..... 588  
228 auTrpValTyrCysGlyAsnGlyThrProAsnGluLeuGlyClyAlaAsn 244  
589 ...CCGGCGACCTGAAACCCAGATCGACAGGTGAGTGCAGCGCAGGT 635  
245 ileProAlaGluPheLeuGluAsnGlyValArgSerSerAsnLeuLysPh 261  
636 TCGTTTCAGGCGCAGTGGCGCGCGCGCGGACGG ..... C 673  
261 eGlnAspAlaTyrAsnAlaAlaGlyGlyHisAsnAlaValPheAsnPh 278  
674 CGCCAGCGCGGCGTGGTGGCTTCCAAAGACGACCAATAAGCAGAAGC 723  
278 toPro. .... 279  
724 AGGAATCGACGAGATCTCGACGAATATTCGTACGCGCGCGTCCAAATAC 773  
280 ...AsnGlyThrHisSerTrp.GluTyrTrpGlyAlaGlnLeuAsnAlaM 295  
774 TCGAGGCGCGGAGGAGCAGCAGCGCGTCTCTCGCAATGGGCTT 823  
295 eLysGlyAspLeuGlnSerSerLeuGlyAla ..... GlyLys 307  
824 CTGACCCGCTAATACGAAAGAAAGACGAGCAAAACATGACAGACGCA 873  
308 LeuAla ..... MetThrGluGlnG 314  
874 GTGGAAATTCGGGGTATCGAGCGCGCGCAAGCGCAATCCAGGAAATG 923  
314 nTrpAsnPheAlaGlyIleGluAlaAlaAlaSerAlaIleGlnGlyAsnV 331  
924 TCACCTCCATTCATTCCTCTCTGACGAGGAGGAGCAGTCCCTGACCAAG 973  
331 alrSerIleHisSerLeuLeuAspGlyGlyLysGlnSerLeuThrLys 347  
974 CTCGAGCGCGCTGGGGGTAGCGGTTCGAGGCGGTACCGAGGTGTCCA 1023  
348 LeuAlaAlaAlaTrpGlySerGlySerGluAlaTyrGlnGlyValG 364  
1024 GCAAAATGGAGCGCCAGCGGTACCGAGCTGAACACCGCGTGCAGAAC 1073  
364 nGlnLysTrpAspAlaThrAlaThrGluLeuAsnAlaLeuGlnAsnL 381  
1074 TGGCGCGGACGATCAGCAGCGGTACGCAATGCTTCGACCGCAAGGC 1123  
381 euAlaArgThrIleSerGluAlaGlyGlnAlaMetAlaSerThrGluGly 397  
1124 AACGTCACCTGGGATGTCGCA 1144  
398 AsnValThrGlyMetPheAla 404

seq\_name: /SIDS1/gcdata/hold-geneseq/geneeq-embl/AA1999.DAT:AA1999

seq\_documentation\_block:

ID\_AA1999 standard; Protein; 404 AA.

XX AC AA1999;

XX DT 06-SEP-1999 (first entry)

XX DE Amino acid sequence of antigen MPT59.

XX KW Immunogenic; Mycobacterium tuberculosis; immune response; infection;  
XX KW tuberculosis; fusion polypeptide; T-cell epitope; ESAT-6; MPT59; TB;  
XX KW pharmaceutical; vaccination; M. africanum; M. bovis; CFP30A; CFP7B;  
XX KW CFP19; CFP27; CFP30A; RDI-ORF; CFP16; CFP19; CFP23;  
XX KW CFP25A; CFP30B; CFP7B.  
XX OS Mycobacterium tuberculosis.

PN W09924577-A1.  
XX 20-MAY-1999.  
XX 08-OCT-1998; 98WO-DK00438.  
XX 01-APR-1998; 98WO-DK00132.  
PR 10-NOV-1997; 97DK-0001277.  
XX 05-JAN-1998; 98US-0070488.  
XX (STAT-) STATENS SERUM INST.  
XX Andersen P, Skjot R;  
XX WPI; 1999-347282/29.  
XX New immunogenic fragment of Mycobacterium tuberculosis  
PT  
XX  
XX Examples; Page 249-250; 265pp; English.  
XX  
CC The invention describes a substantially pure immunogenic polypeptide  
CC fragment (I) from Mycobacterium tuberculosis that is able to evoke a  
CC protective immune response against infections by mycobacteria belonging  
CC to the tuberculosis complex. The invention provides a (1) fusion  
CC polypeptide comprising at least one polypeptide fragment (I) and at least  
CC one fusion partner; (2) a fusion polypeptide fragment comprising a T-cell  
CC epitope from M. tuberculosis protein ESAT-6, or MPT59 and a second  
CC different amino acid sequence from M. tuberculosis, and/or including a  
CC sequence which protects the first amino acid sequence from in vivo  
CC degradation or post-translational processing; (3) a nucleic acid fragment  
CC that encodes the above polypeptides. The polypeptides and nucleic acid  
CC are useful as pharmaceuticals, for diagnosis of and as antigens for  
CC vaccination against TB caused by Mycobacterium tuberculosis, africanum or  
CC bovis. The polypeptides are also useful for diagnosing ongoing or  
CC previous sensitization in an animal with bacteria belonging to the  
CC tuberculosis complex. The invention also describes the use of CFP7A or  
CC CFP30A or a T-cell epitope of for the induction of a strong immune  
CC response in a mammal; use of CFP7B, CFP19 or MPT59-ESAT6 or a T-cell  
CC epitope of for diagnosis of TB in a mammal by performing a DTH type skin  
CC test; use of CFP27, CFP30A, RDI-ORF2, RDI-ORF3, MPT59-ESAT6,  
CC ESAT6-MPT59, CFP10A, CFP16, CFP19, CFP23, CFP25A, CFP30B, CFP7B or a T-  
CC cell epitope of for the preparation of an immunological composition; and  
CC for the preparation of a subunit vaccine.  
XX  
XX Sequence 404 AA;

alignment\_scores:

Quality: 487.50 Length: 391

Ratio: 2.579 Gaps: 13

Percent Similarity: 48.338 Percent Identity: 36.829

alignment\_block:

US-09-462-480-1 x AA1999

Align seg 1/1 to: AA1999 from: 1 to: 404

328 CAATCCGCGCGCTCCACCGAGCGCGGTCTGGTCGCGCGCGCGCGCGCGC 377

|||||

48 GlnSerGlyGlyAsnAsnSerProAlaValTyrLeu.....LeuAs 61

378 GCAGGAGCGTGAAGAAGACGAGGAGCGAGTGGGACGAAGAG..... 420

|||

61 pGlyLeuArgAlaGlnAspAspTyrAsnGlyTrpAspIleAsnThrProA 78

420 .....

78 laPheGluTrpTyrGlnSerGlyLeuSerIleValMetProValGly 94

421 .....

95 GlyGlnSerSerPheTyrSerAspTrpTyrSerProAlaCysGlyLysAl 111

442 .....ATGACAACAGACTTCCCG. 459  
111 aGlyCysGlnThrTyrlsYsrpGluThrPheLeuThrSerGluLeuProG 128  
459 ..... 459  
128 lnTrpLeuSerAlaAsnArgAlaVallysProThrGlySerAlaAlaile 144  
460 ...GCCACCGCGCGGAGACATTCACACATTTTGGCGAGGAGGTAAA 506  
145 GlyLeuSerMetAlaGlySerSerAlaMetIleLeuAlaAlaTyrHisPr 161  
507 GAGAGAA..... 513  
161 oGlnGlnPheIleTyrrAlaGlySerLeuSerAlaLeuLeuAspProSerG 178  
513 ..... 513  
178 lnGlyMetGlyProSerLeuIleGlyLeuAlaMetGlyAspAlaGlyGly 194  
514 .....AGTAGTCCAGCATGGCAGAG 533  
195 TyrlsAlaAlaaspMetTrpGlyProSerSerAspProAlaTrpGluAr 211  
534 ATGAGAGCG.....ATGCCG.....CTACCCCTCGGCGAGGAGGC 568  
211 gAsnAspProThrGlnGlnIleProLysLeuValAlaAsnThrArgL 228  
569 AGTAATTTGAGCGGATCT..... 588  
228 euTrpValTyrcysGlyAsnGlyThrProAsnGluLeuGlyGlyAlaAsn 244  
589 ...CCGCGGACCTGAAACCCAGATCCAGATCCAGGTTGGAGTCGACGCGAGGT 635  
245 IleProAlaGluPheLeuGluAsnPhelArgSerSerAsnLeuLysPh 261  
636 TCGTTCGAGGCGCAGTGGCGCGCGCGCGGCGGACGG.....C 673  
261 eGlnAspAlaTyrrAsnAlaAlaGlyGlyHisAsnAlaValPheAsnPh 278  
674 CGCCCGCGCGCGTGTGGCTTCCAAAGACGAGCCAAATAGCAGAGC 723  
278 roPro..... 279  
724 AGGAACTCGACGAGATCTCGACGAATATTCGTACGCGCGCGTCCAATAC 773  
280 ...AsnGlyThrHisSerTrp.GluTyrrpGlyAlaGlnLeuAsnAlaM 295  
774 TCGAGGCGCGCAGGAGGACGACGAGCGCGCTGTCTCGCAATGGCGTT 823  
295 etLysGlyAspLeuGlnSerSerLeuGlyAla.....GlyLys 307  
824 CTGACCGCGCTAATACGAAGAAACGAGCAAAACATGACAGAGCAGCA 873  
308 LeuAla.....MetThrGluGlnG 314  
874 GTGAATTTTCGGGTATCGAGGCGCGGCAAGCGCAATCCAGGAAATG 923  
314 nTrpAsnPheAlaGlyIleGluAlaAlaAlaSerAlaIleGlnGlyAsnV 331  
924 TCAGCTCATTCATTCCTCTCTGACGAGGGAAGCAGTCCTCGACCAAG 973  
331 alThrSerIleHisSerLeuLeuAspGluGlyLysGlnSerLeuThrLys 347  
974 CTCGACGCGCGCTGGCGGTAGCGGTTCGGAGCGGTACAGGCGTGTCCA 1023  
348 LeuAlaAlaAlaTrpGlyGlySerGlySerGluAlaTyrrGlnGlyValG 364  
1024 GCAAAATGGGACCGCGCTACCGAGCTGAACAACGCGCTCGAGAAC 1073  
364 nGlnLysTrpaspAlaThrAlaThrGluLeuAsnAlaLeuGlnAsnL 381  
1074 TGGCGCGGACGATCAGCGAAGCGCGTCAGGCAATGGCTTCGACCGAAGGC 1123

|||||  
381 euAlaArgThrIleSerGluAlaGlyGlnAlaMetAlaSerThrGluGly 397  
1124 AACGTCACTGGGATGTCGCA 1144  
398 AsnValThrGlyMetPheAla 404  
seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.AAW81746  
seq\_documentation\_block:  
ID AAW81746 standard; Protein; 802 AA.  
XX AAW81746;  
XX AC  
XX 27-JAN-1999 (first entry)  
XX DE M. tuberculosis fusion protein TbF-2.  
XX KW Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;  
XX KW vaccine; pharmaceutical; infection; diagnosis.  
XX OS Synthetic.  
XX OS Mycobacterium tuberculosis.  
XX PN W09816646-A2.  
XX PD 23-APR-1998.  
XX PF 07-OCT-1997; 97WO-US18293.  
XX PR 13-MAR-1997; 97US-0818112.  
XX PR 11-OCT-1996; 96US-0730510.  
XX PA (CORI-) CORIXA CORP.  
XX PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;  
PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;  
XX WPI: 1998-261042/23.  
XX DR N-PSDB; AAV64567.  
XX CC Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used  
PT to develop products for the detection of M. tuberculosis infection  
PT and for diagnosis, treatment and prevention of tuberculosis  
XX PS Disclosure; Page 208-211; 230pp; English.  
XX CC This sequence represents the fusion protein TbF-2 which is composed of  
CC immunogenic polypeptides from Mycobacterium tuberculosis (MT). This  
CC protein is used in a method for inducing protective immunity against  
CC tuberculosis (TB). This sequence can be formulated into vaccines  
CC and/or pharmaceutical compositions for immunising against  
CC M. tuberculosis infection or may be used for the diagnosis of TB.  
XX SQ Sequence 802 AA;

alignment\_scores:  
Quality: 487.00 Length: 103  
Ratio: 4.822 Gaps: 0  
Percent Similarity: 98.058 Percent Identity: 97.087  
alignment\_block:  
US-09-462-480-1 x AAW81746 ..  
Align seg 1/1 to: AAW81746 from: 1 to: 802  
522 AGCATGCGAGATGAAGACGATCCCGCTACCTCGGCGAGGAGCAGG 571  
423 SerSerAlaGluMetLysThrAspAlaThrLeuAlaGlnGluAlaG 439  
572 TAATTTTCAGCGGATCTCCGCGCGACCTGAAACCCAGATCGACCGGTGG 621  
|||||

439 yAsnPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValG 456  
622 AGTCAGCGCAGGTGCTGTCAGGCGCCAGTGGCGCGCGCGGGGACG 671  
|||||  
456 luSerThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThr 472  
672 GCCGCCAGCGCGGGTGGTGGCTTCCAAAGCAGCAGCCAAATAGCAGAA 721  
|||||  
473 AlaAlaGlnAlaAlaValValArgPheGlnGluAlaAlaAsnLysGlnLy 489  
722 GCAGGAACCTGCAGCAGATCTCGAGCAATATTCTGACGCGCGCTCCAAT 771  
|||||  
489 sGlnGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnT 506  
772 ACTCGAGCGCGCAGCAGCAGCAGCGCTGTCTCGCAAAATGGCG 821  
|||||  
506 yrSerArgAlaAspGluGlnGlnGlnAlaLeuSerSerGlnMetGly 522  
822 TTCTGACCC 830  
||| |||  
523 PheValPro 525

seq\_name: /SDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.AAW64379.

seq\_documentation\_block:

ID AAW64379 standard; Protein; 802 AA.

XX AAW64379;

XX AC

XX DT 09-NOV-1998 (first entry)

DE Mycobacterium antigen Tbf2 protein fusion.

XX Tuberculosis; infection; diagnosis; 38 kDa antigen; Tbra3; DPEP;

XX Tb38-1; Tbf-2.

XX Mycobacterium tuberculosis.

OS Synthetic.

XX WO9816645-A2.

XX 23-APR-1998.

XX 07-OCT-1997; 97WO-US18214.

XX 13-MAR-1997; 97US-0818111.

XX 11-OCT-1996; 96US-0729622.

XX (CORI-) CORIXA CORP.

XX Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;

PI Reed SG, Skelky YAW, Twardzik DR, Vedvick TS;

XX WPI; 1998-251292/22.

XX N-PSDB; AAV55801.

XX New isolated Mycobacterium tuberculosis polypeptides and DNA - used

PT to develop products for the detection of M. tuberculosis infection

PT and diagnosis of tuberculosis

XX Example 7; Page 223-226; 250pp; English.

XX This polypeptide comprises a fusion protein, designated Tbf-2,

CC composed of Mycobacterium tuberculosis antigens Tbra3 (see AAW64295),

CC 38 kDa antigen (see AAW64364), Tb38-1 (see AAW64321) and DPEP (see

CC AAW64322). It was produced by PCR amplification (see AAV44450-57) of

CC the appropriate antigen DNA sequences, cloning into an expression

CC vector, and expression in E. coli. Tbf-2 can be used for

CC serodiagnosis of tuberculosis. The invention relates to

CC comparisons and methods for diagnosing tuberculosis. It provides

CC polypeptides (see AAW64291-W64379) comprising antigenic or

CC immunogenic portions of M. tuberculosis antigens, or fusion proteins,

CC DNA sequences encoding such polypeptides, recombinant expression

CC vectors and host cells. Also claimed are methods and diagnostic

CC kits for detecting M. tuberculosis infection in a patient.

XX SQ Sequence 802 AA;

alignment\_scores:

Quality: 487.00 Length: 103

Ratio: 4.822 Gaps: 0

Percent Similarity: 98.058 Percent Identity: 97.087

alignment\_block:

US-09-462-480-1 x AAW64379

Align seg 1/1 to: AAW64379 from: 1 to: 802

522 AGTCAGCAGAGATCAAGACCGATGCCGCTACCTCGGGCAGGAGGACG 571

||| |||

423 SerSerAlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaG 439

|||||

572 TAATTTCGAGCGGATCTCCGGCGACCTGAAACCCAGATCGACCAAGTGG 621

|||||

439 yAsnPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValG 456

|||||

622 AGTCAGCGCAGGTTCGTTGTCAGGCGCCAGTGGCGCGCGCGGGGACG 671

|||||

456 luSerThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThr 472

|||||

672 GCCGCCAGCGCGGGTGGTGGCTTCCAAAGCAGCAGCCAAATAGCAGAA 721

|||||

473 AlaAlaGlnAlaAlaValValArgPheGlnGluAlaAlaAsnLysGlnLy 489

|||||

722 GCAGGAACCTGCAGCAGATCTCGAGCAATATTCTGACGCGCGCTCCAAT 771

|||||

489 sGlnGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnT 506

|||||

772 ACTCGAGCGCGCAGCAGGAGCAGCAGCGCTGTCTCGCAAAATGGCG 821

|||||

506 yrSerArgAlaAspGluGlnGlnGlnAlaLeuSerSerGlnMetGly 522

|||||

822 TTCTGACCC 830

||| |||

523 PheValPro 525



OM of: US-09-462-480-1 to: Issued\_Patents\_AA:\* out\_format : pfs  
Date: Jul 22, 2002 1:24 AM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh  
-Q/cgn2\_1/USPTO.spool/US09462480/runat\_18072002.164418.19425/app\_query.fasta\_1.2850  
-DB=Issued\_Patents\_AA -OPMT=fastan -SUFFIX=rai -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOPEXT=0.000  
-GAPOP=4.500 -GAPEXT=0.050 -GAPOP=10.000 -XGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09462480\_CGNI\_1\_18 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-462-480-1  
Query length: 1277  
Database: Issued\_Patents\_AA:\*  
Database sequences: 231628  
Database length: 24425594  
Search time (sec): 53.850000

score\_list:

Sequence	Strd Orig	zScore	Escore	Len	Documentation
/cgn2_6/ptodata/2/1aa/6B_COMB.pcp:US-08-818-112-114		713.00	1026.33	368	1.0e-50
/cgn2_6/ptodata/2/1aa/6B_COMB.pcp:US-08-818-111-109		713.00	1026.33	368	3.0e-50
/cgn2_6/ptodata/2/1aa/6B_COMB.pcp:US-09-056-556-114		713.00	1026.33	368	3.0e-50
/cgn2_6/ptodata/2/1aa/6B_COMB.pcp:US-09-056-556-214		487.00	687.90	802	9.7e-32
/cgn2_6/ptodata/2/1aa/6B_COMB.pcp:US-08-818-112-115		486.00	708.68	100	5.4e-32
/cgn2_6/ptodata/2/1aa/6B_COMB.pcp:US-08-818-111-110		486.00	708.68	100	5.4e-32
/cgn2_6/ptodata/2/1aa/6B_COMB.pcp:US-09-056-556-115		486.00	708.68	100	5.4e-32
/cgn2_6/ptodata/2/1aa/6B_COMB.pcp:US-08-456-640-2		479.00	699.01	95	2.0e-31
/cgn2_6/ptodata/2/1aa/6B_COMB.pcp:US-08-818-112-88		462.00	674.18	95	4.8e-30
/cgn2_6/ptodata/2/1aa/6B_COMB.pcp:US-08-818-111-99		462.00	674.18	95	4.8e-30
/cgn2_6/ptodata/2/1aa/6B_COMB.pcp:US-09-056-556-88		462.00	674.18	95	4.8e-30
/cgn2_6/ptodata/2/1aa/6B_COMB.pcp:US-08-818-112-117		392.00	573.77	80	2.2e-24
/cgn2_6/ptodata/2/1aa/6B_COMB.pcp:US-08-818-111-112		392.00	573.77	80	2.2e-24
/cgn2_6/ptodata/2/1aa/6B_COMB.pcp:US-09-056-556-117		392.00	573.77	80	2.2e-24
/cgn2_6/ptodata/2/1aa/6B_COMB.pcp:US-08-818-112-104		259.00	384.32	51	1.2e-13
/cgn2_6/ptodata/2/1aa/6B_COMB.pcp:US-08-818-111-99		259.00	384.32	51	1.2e-13
/cgn2_6/ptodata/2/1aa/6B_COMB.pcp:US-09-056-556-104		259.00	384.32	51	1.2e-13
/cgn2_6/ptodata/2/1aa/6B_COMB.pcp:US-09-199-637A-277		209.50	286.46	558	3.2e-09
/cgn2_6/ptodata/2/1aa/6B_COMB.pcp:US-09-413-814-78		181.00	239.94	882	7.9e-07
/cgn2_6/ptodata/2/1aa/6B_COMB.pcp:US-07-945-283-2		180.00	238.48	882	9.6e-07
/cgn2_6/ptodata/2/1aa/6B_COMB.pcp:US-09-041-886-23		165.00	213.42	1185	1.8e-05
/cgn2_6/ptodata/2/1aa/6B_COMB.pcp:US-09-413-814-79		164.00	211.71	1213	2.2e-05
/cgn2_6/ptodata/2/1aa/6B_COMB.pcp:US-08-483-533-41		160.50	219.72	355	2.6e-05
/cgn2_6/ptodata/2/1aa/6B_COMB.pcp:US-09-283-471A-41		160.50	219.72	355	2.6e-05
/cgn2_6/ptodata/2/1aa/6B_COMB.pcp:PCT-US91-06532-3		160.50	219.72	355	2.6e-05
/cgn2_6/ptodata/2/1aa/6B_COMB.pcp:US-07-928-611-13		159.00	220.57	267	3.1e-05
/cgn2_6/ptodata/2/1aa/6B_COMB.pcp:US-08-487-811A-13		159.00	220.57	267	3.1e-05
/cgn2_6/ptodata/2/1aa/6B_COMB.pcp:US-09-060-694-13		159.00	220.57	267	3.1e-05
/cgn2_6/ptodata/2/1aa/6B_COMB.pcp:PCT-US93-07370-13		159.00	220.57	267	3.1e-05
/cgn2_6/ptodata/2/1aa/6B_COMB.pcp:US-08-056-051-6		159.00	214.60	467	3.9e-05
/cgn2_6/ptodata/2/1aa/6B_COMB.pcp:US-07-928-611-22		159.00	214.60	467	3.9e-05
/cgn2_6/ptodata/2/1aa/6B_COMB.pcp:US-08-487-811A-22		159.00	214.60	467	3.9e-05
/cgn2_6/ptodata/2/1aa/6B_COMB.pcp:US-09-060-694-22		159.00	214.60	467	3.9e-05
/cgn2_6/ptodata/2/1aa/6B_COMB.pcp:PCT-US93-07370-22		159.00	214.60	467	3.9e-05
/cgn2_6/ptodata/2/1aa/6B_COMB.pcp:US-08-931-820-2		156.00	201.83	1024	9.1e-05
/cgn2_6/ptodata/2/1aa/6B_COMB.pcp:US-08-963-825-19		156.00	198.75	1366	0.0001
/cgn2_6/ptodata/2/1aa/6B_COMB.pcp:US-09-570-573-19		156.00	198.75	1366	0.0001
/cgn2_6/ptodata/2/1aa/6B_COMB.pcp:US-09-548-608-19		156.00	198.75	1366	0.0001
/cgn2_6/ptodata/2/1aa/6B_COMB.pcp:US-08-316-650-12		155.00	196.71	1442	0.0001

/cgn2\_6/ptodata/2/1aa/6B\_COMB.pcp:PCT-US95-02251-12 + 155.00 196.71 0.0001 1213  
/cgn2\_6/ptodata/2/1aa/6B\_COMB.pcp:US-09-413-814-79 - 153.50 196.37 0.0002 747  
/cgn2\_6/ptodata/2/1aa/6A\_COMB.pcp:US-09-034-177-3 + 152.00 199.36 0.0002 1078  
/cgn2\_6/ptodata/2/1aa/6A\_COMB.pcp:US-08-963-825-21 + 152.00 195.44 0.0002 1078  
/cgn2\_6/ptodata/2/1aa/6B\_COMB.pcp:US-09-570-573-21 + 152.00 195.44 0.0002 1078  
seq\_name: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pcp:US-08-818-112-114  
seq\_documentation\_block:  
; Sequence 114, Application US/08818112  
; Patent No. 6290969  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Campos-Neto, Antonio  
; APPLICANT: Houghton, Raymond  
; APPLICANT: Vedvick, Thomas S.  
; APPLICANT: Twardzik, Daniel R.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY  
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS  
; NUMBER OF SEQUENCES: 153  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/818,112  
; FILING DATE: 13-MAR-1997  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.411C6  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 114:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 368 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-818-112-114  
alignment\_scores:  
Quality: 713.00 Length: 143  
Ratio: 5.057 Gaps: 0  
Percent Similarity: 98.601 Percent Identity: 96.503  
alignment\_block:  
US-09-462-480-1 x US-08-818-112-114 ..  
Align seg 1/1 to: US-08-818-112-114 from: 1 to: 368.  
1 CTCAGCAGGTGACGTGTTGTTACGCCAGGTGGCGGCGCCGCGCGG 50  
|||||  
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGly 242  
51 CAACCCAGCCGACGAGGAGCGCGGCGGCGGCTGCTCGGCGCCAGTC 100  
|||||  
242 YasnProAlaAspGluGluAlaGlnMetGlyLeuLeuGlyThrSerP 259  
101 CGCTGTGCAACCATCCGCTGGCTGGTGGATCAGGCCCGCGGCGCG 150



```
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 114:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 368 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-056-556-114

alignment_scores:
  Quality: 713.00      Length: 143
  Ratio: 5.057         Gaps: 0
  Percent Similarity: 98.601  Percent Identity: 96.503

alignment_block:
  US-09-462-480-1 x US-09-056-556-114 ..
  Align seg 1/1 to: US-09-056-556-114 from: 1 to: 368

1 CTGCAGCAGGTGACGTCGCTGCTTCTCAGCCAGGTGGCGCGCGCGCGG 50
  |||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGlyGly 242
  |||

51 CAACCCAGCCGACGAGGAGCGCGGCGGCGGCGGCGGCGGCGGCGG 100
  |||
242 YAsnProAlaAspGluGluAlaAlaGlnMetGlyLeuLeuGlyThrSerP 259
  |||

101 CGTGTGCAACCATCCGCTGCTGTGTGATCAGGCCCGCGCGCGCGG 150
  |||
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275
  |||

151 GGCCTGCTGCGCGGCGGAGTCTACCTGCGCGGCGGCGGCGGCGG 200
  |||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292
  |||

201 CACGCCGCTGATGTCTCAGCTGATCGAAAGCGGTTGCCCGCTCGGTGA 250
  |||
292 gThrProLeuMetSerGlnLeuGlyGlyProValAlaProSerValM 309
  |||

251 TCCCGCGCGGTGTCGCGGATCGTGGTGACGGGTGGCGCGGCGGCG 300
  |||
309 etProAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
  |||

301 GGTCCGGGAGCGATGGCGGCGGTCGCAATCCGCGGCGGCTCCACGCCC 350
  |||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
  |||

351 GGGTGTGTCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 400
  |||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAspG 359
  |||

401 AGGAGCGACTGGGACGAAGAGCGAGCTGG 429
  |||
359 luAspAspTrpAspGluGluAspAspTrp 368
  |||

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-056-556-214

seq_documentation_block:
; Sequence 214, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, David C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
```

TREATM

```
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 214:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 802 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-056-556-214

alignment_scores:
  Quality: 487.00      Length: 103
  Ratio: 4.822         Gaps: 0
  Percent Similarity: 98.058  Percent Identity: 97.087

alignment_block:
  US-09-462-480-1 x US-09-056-556-214 ..
  Align seg 1/1 to: US-09-056-556-214 from: 1 to: 802

522 AGCATGCGCAGATGAAGACCGATCCGCTACCCCTCGGCGGCGGCGG 571
  |||
423 SerSerAlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGl 439
  |||

572 TAATTTGAGCGGATCTCCGCGACCTGAAACCCAGATCGACCGGTGG 621
  |||
439 yAsnPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValG 456
  |||

622 AGTCACGCGCAGGTTGCTGTCAGGCGGCGGCGGCGGCGGCGGCGG 671
  |||
456 luSerThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThr 472
  |||

672 GCGCGCCGCGCGGCTGCTGCTTCCAAAGACGAGCAATAAGCAGAA 721
  |||
473 AlaAlaGlnAlaAlaValAlaArgPheGlnGluAlaAlaAsnLysGlnLy 489
  |||

722 GCAGGAATCGACGAGATCTCGACGAATATTGTCAGGCGGCGGCTCAAT 771
  |||
489 sgGlnGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnT 506
  |||

772 ACTCAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 821
  |||
506 yrSerArgAlaAspGluGlnGlnGlnAlaLeuSerSerGlnMetGly 522
  |||

822 TTCTGACCC 830
  |||
523 PheValPro 525
  |||

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-818-112-115
```

seq\_documentation\_block:  
; Sequence 115, Application US/08818112  
; Patent No. 6290969  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Dillon, David C.  
; APPLICANT: Campos-Neto, Antonio  
; APPLICANT: Houghton, Raymond  
; APPLICANT: Vedvick, Thomas S.  
; APPLICANT: Twardzik, Daniel R.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY  
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS  
; NUMBER OF SEQUENCES: 153  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED AND BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/818,112  
; FILING DATE: 13-MAR-1997  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.411C6  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 115:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 100 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
US-08-818-112-115

alignment\_scores:  
Quality: 486.00 Length: 100  
Ratio: 4.860 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 99.000

alignment\_block:  
US-09-462-480-1 x US-08-818-112-115 ..  
Align seg 1/1 to: US-08-818-112-115 from: 1 to: 100

```
525 ATGGCAGAGATGACGACCGATGCCCTACCTCGCGCAGGAGGTA 574
|||||
1 MetAlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAs 17
575 TTTCAGCGGATCTCCGGCGACCTGAAACCCAGATCGACAGGTGGAGT 624
|||||
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGlu 34
625 CGACGCGAGGTTCGTTGTCAGGCGCCAGTGGCGCGCGGCGGCGCC 674
|||||
34 erThAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50
675 GCCACGCGCGGTTGGTGGCTTCCAAAGCAGCAGCAATAAGCAGCA 724
|||||
51 AlaGlnAlaAlaValAlaArgPheGlnGluAlaAlaAsnLysGlnLysG 67
725 GGAACCTGACGAGATCTCGACGAATATTCGTGACGCGCGCGGTCCAATACT 774
```

|||||  
67 nGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnIle 84  
775 CGAGGCGCCGACGAGGAGCAGCAGCGCTGTCTCCGCAAAATGGGCTTC 824  
|||||  
84 erArgAlaAspGluGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 100  
seq\_name: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pap:US-08-818-111-110  
seq\_documentation\_block:  
; Sequence 110, Application US/08818111  
; Patent No. 633852  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Dillon, David C.  
; APPLICANT: Campos-Neto, Antonio  
; APPLICANT: Houghton, Raymond  
; APPLICANT: Vedvick, Thomas S.  
; APPLICANT: Twardzik, Daniel R.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED AND BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/818,111  
; FILING DATE: 13-MAR-1997  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.417C6  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 110:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 100 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
US-08-818-111-110

alignment\_scores:  
Quality: 486.00 Length: 100  
Ratio: 4.860 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 99.000

alignment\_block:  
US-09-462-480-1 x US-08-818-111-110 ..  
Align seg 1/1 to: US-08-818-111-110 from: 1 to: 100

```
525 ATGGCAGAGATGACGACCGATGCCCTACCTCGCGCAGGAGGTA 574
|||||
1 MetAlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAs 17
575 TTTCAGCGGATCTCCGGCGACCTGAAACCCAGATCGACAGGTGGAGT 624
|||||
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGlu 34
625 CGACGCGAGGTTCGTTGTCAGGCGCCAGTGGCGCGCGGCGGCGCC 674
```

|||||  
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThraAla 50  
675 GCCAGGCGCGGTGGTGGCTTCCAGAGCAGCAGCAATAGCAGAGCA 724  
|||||  
51 AlaGlnAlaAlaValAlaValArgPheGlnGlnAlaAlaAsnLysGlnLysG1 67  
725 GGAACCTGACGAGATCTCGAGCAATATTCGTCAGGCGCGCTCCAACTACT 774  
|||||  
67 nGluLeuAspGluLeuSerThrAsnIleArgGlnAlaGlyValGlnTrp 84  
775 CGAGGCGCGAGCAGCAGCAGCAGCGCTGCTCTCCGAATAGGCTTC 824  
|||||  
84 erArgAlaAspGluGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 100

seq\_name: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep:US-09-056-556-115

seq\_documentation\_block:  
; Sequence 115, Application US/09056556  
; Patent No. 6350456  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Tasir A.W.  
; APPLICANT: Dillon, David C.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND  
; NUMBER OF SEQUENCES: 241  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED AND BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/056,556  
; FILING DATE: 07-APR-1998  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.457  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 115:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 100 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; US-09-056-556-115

alignment\_scores:  
Quality: 486.00 Length: 100  
Ratio: 4.860 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 99.000

alignment\_block:

US-09-462-480-1 x US-09-056-556-115 ..

Align seg 1/1 to: US-09-056-556-115 from: 1 to: 100

525 ATGGCAGAGATGAACACCGATCGCGCTACCTCGGCGAGGAGGAGTAA 574  
|||||  
1 MetAlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAs 17  
575 TTTCAGCGGATCTCCGCGCGACCTGAAACCCAGATCGACCGAGTGGAGT 624

|||||  
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGluS 34  
625 CGAGGCGAGGTTCCTTCAGGCGCAGTGGCGCGCGGGGGGAGCGGCC 674  
|||||  
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThraAla 50  
675 GCCAGGCGCGGTGGTGGCTTCCAGAGCAGCAGCAATAGCAGAGCA 724  
|||||  
51 AlaGlnAlaAlaValAlaValArgPheGlnGlnAlaAlaAsnLysGlnLysG1 67  
725 GGAACCTGACGAGATCTCGAGCAATATTCGTCAGGCGCGCTCCAACTACT 774  
|||||  
67 nGluLeuAspGluLeuSerThrAsnIleArgGlnAlaGlyValGlnTrp 84  
775 CGAGGCGCGAGCAGCAGCAGCAGCGCTGCTCTCCGAATAGGCTTC 824  
|||||  
84 erArgAlaAspGluGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 100

seq\_name: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep:US-08-465-640-2

seq\_documentation\_block:  
; Sequence 2, Application US/08465640  
; Patent No. 5955077  
; GENERAL INFORMATION:  
; APPLICANT: ANDERSEN, Peter  
; APPLICANT: ANDERSEN, Ase Bengaard  
; APPLICANT: HASLOV, Kaare  
; APPLICANT: SORENSEN, Anne Lund  
; TITLE OF INVENTION: TUBERCULOSIS VACCINE  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Browdy and Neimark  
; STREET: 419 Seventh Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/465,640  
; FILING DATE: 05-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/123,182  
; FILING DATE: 20-SEP-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/DK94/00273  
; FILING DATE: 01-JUL-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: COOPER, IVER P  
; REGISTRATION NUMBER: 28,005  
; REFERENCE/DOCKET NUMBER: ANDERSEN-3A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-628-5197  
; TELEFAX: 202-737-3528  
; TELEX: 248633  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 95 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-465-640-2

alignment\_scores:  
Quality: 479.00 Length: 95  
Ratio: 5.042 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

```
alignment_block:
US-09-462-480-1 x US-08-465-640-2
Align seg 1/1 to: US-08-465-640-2 from: 1 to: 95

860 ATGACAGACGACAGTGGAAATTCGCGGTATCGAGCGCGGCAAGCGC 909
|||||
1 MetThrGluGlnGlnTrpAsnPheAlaGlyIleGluAlaAlaSerAl 17

910 AATCCAGGGAATGTCAGCTCATTCATTCCTCTGACGAGGGAAGC 959
|||||
17 aileGlnGlyAsnValThrSerIleHisSerLeuLeuAspGluGlyysG 34

960 AGTCCTGACCAAGCTCGACGCGCTGGGGCGGTAGCGTTTCGAGCGC 1009
|||||
34 InSerLeuThrLysLeuAlaAlaAlaAlaTrpGlySerGlySerGluAla 50

1010 TACGAGGCTGCCAGCAAAATGGAGCCACCGCTACCGAGCTGAACAA 1059
|||||
51 TyrGlnGlyValGlnGlnLysTrpAspAlaThrAlaThrGluLeuAsnAs 67

1060 CGCGCTGCAAGAACCTGGCGCGGAGATCAGCAAGCGCGGTTCAGCAATGG 1109
|||||
67 nAlaLeuGlnAsnLeuAlaArgThrIleSerGluAlaGlyGlnAlaMeta 84

1110 CTTGACCGAAGCAAGCTCACTGGGATGTCGCA 1144
|||||
84 laSerThrGluGlyAsnValThrGlyMetPheAla 95

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-818-112-88

seq_documentation_block:
; Sequence 88, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 88:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 95 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-112-88

alignment_scores:
Quality: 462.00 Length: 95
Ratio: 4.863 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.947

alignment_block:
US-09-462-480-1 x US-08-818-112-88
Align seg 1/1 to: US-08-818-112-88 from: 1 to: 95

540 ACCGATCCCGCTACCTCGGCGAGGAGCGAGTAAATTCGAGCGGATCTC 589
|||||
1 ThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAsnPheGluArgIleSe 17

590 CGCGGACCTGAAACCCAGATCGACAGTGGAGTGGAGTGGAGCGCGGTCT 639
|||||
17 rGlyAspLeuLysThrGlnIleAspGlnValGluSerThrAlaGlySerL 34

640 TGCAGGCGCAGTGGCGCGCGCGGCGGCGCGCGCGCGCGCGCGGTG 689
|||||
34 euGlnGlyGlnTrpArgGlyAlaAlaGlyThrAlaAlaGlnAlaAlaVal 50

690 GTGCGCTTCCNAGAGCAGCCCAATAAGCAGAGGAGGAGGAGGAGGAT 739
|||||
51 ValArgPheGlnGluAlaAlaAsnLysGlnLysGlnGluLeuAspGluIl 67

740 CTCGACGAATATTCGTGAGCGCGCGCTCTCTCGCAATGCGCTTC 824
|||||
84 LuGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 95

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-818-111-89

seq_documentation_block:
; Sequence 89, Application US/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,111
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 89:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 95 amino acids
; TYPE: amino acid
```

```

; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELECOMMUNICATION INFORMATION:
; NAME: Maki, David J.
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 89:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 95 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-111-89

alignment_scores:
    Quality: 462.00      Length: 95
    Ratio: 4.863        Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 98.947

alignment_block:
US-09-462-480-1 x US-08-818-111-89 ..
Align seg 1/1 to: US-08-818-111-89 from: 1 to: 95

540 ACCGATGCCGTACCTCGGCGAGGAGCGAGTAAATTCGAGCGGATCTC 589
|||||
1 ThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAsnPheGluArgIleSe 17

590 CGGCGACCTGAAACCCAGATCGACAGGTGGAGTCGACGCGGATTCGT 639
|||||
17 rGlyAspLeuLysThrGlnIleAspGlnValGluSerThrAlaGlySerL 34

640 TGCAGGCGCAGTGGCGGCGCGCGGAGCGCGCCGCGCGCGGTG 689
|||||
34 euGlnGlyGlnTrpArgGlyAlaAlaGlyThrAlaAlaGlnAlaAlaVal 50

690 GTGCGCTTCCAAGAGCAGCAATAAGCAGAGCAGCAACTCGACGAGAT 739
|||||
51 ValArgPheGlnGluAlaAlaAsnLysGlnLysGlnGluLeuAspGluIl 67

740 CTCGACGAATATTCGTGCGCGCGTCCCAATATGGCGCTTC 824
|||||
790 AGCAGCAGCAGCGCGTGTCTCGCAAAATGGCGCTTC 824
|||||
84 luGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 95

seq_name: /cgn2_6/ptodata/2/iaa/68_COMB.pep:US-09-056-556-88

seq_documentation_block:
; Sequence 88, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998

```





STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/056,556  
FILING DATE: 07-APR-1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.457  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 117:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 80 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-056-556-117

alignment\_scores:  
Quality: 392.00 Length: 80  
Ratio: 4.900 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-462-480-1 x US-09-056-556-117 ..

Align seg 1/1 to: US-09-056-556-117 from: 1 to: 80

585 ATCTCCGGCGACCTCAAAACCCAGATCCAGCAGGTGGAGTCGACGGCAGG 634  
|||||  
1 IleSerGlyAspLeuLysThrGlnIleAspGlnValGluSerThrAlaG1 17  
635 TTCGTTGCGAGCGCCAGTCGCGCGCGCGGCGGACGCCGCCAGCGCG 684  
|||||  
17 ySerLeuGlnGlyGlnTrpA-gGlyAlaAlaGlyThrAlaAlaGlnAlaA 34  
685 CGTGGTGGCTTCCAAAGACGAGCCCAATAAGCAGACGAGCAACTCGAC 734  
|||||  
34 laValValArgPheGlnGluAlaAlaAsnLysGlnLysGlnGluLeuAsp 50  
735 GAGATCTCGAGCAATATTCGTGACGCGCGGCTCCCAATACTCGAGGCGGA 784  
|||||  
51 GluIleSerThrAsnIleArgGlnAlaGlyValGlnTyrSerArgAlaAs 67  
785 CGAGGAGCAGCAGCAGCGCTGTCTCTCGCAATGGCTTC 824  
|||||  
67 pGluGluGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 80

seq\_name: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep:US-08-818-112-104

seq\_documentation\_block:

Sequence 104, Application US/08818112  
Patent No. 6290969

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.  
APPLICANT: Skeiky, Yasir A.W.  
APPLICANT: Dillon, Davin C.  
APPLICANT: Campos-Neto, Antonio  
APPLICANT: Houghton, Raymond  
APPLICANT: Vedvick, Thomas S.  
APPLICANT: Twardzik, Daniel R.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY  
TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS  
NUMBER OF SEQUENCES: 153  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED AND BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/818,112  
FILING DATE: 13-MAR-1997  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.411C6  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 104:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 51 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-818-112-104

alignment\_scores:  
Quality: 259.00 Length: 51  
Ratio: 5.078 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-462-480-1 x US-08-818-112-104 ..

Align seg 1/1 to: US-08-818-112-104 from: 1 to: 51

860 ATGACAGACGACGACTGGAATTTCGCGGTATCGAGCGCGCAAGCGC 909  
|||||  
1 MetThrGluGlnGlnTrpAsnPheAlaGlyIleGluAlaAlaAlaSerAl 17  
910 AATCCAGGGAATGTACGTCCATTTCCTCTTGCAGGAGGGAAGC 959  
|||||  
17 alIeGlnGlyAsnValThrSerIleHisSerLeuLeuAspGluGlyLysG 34  
960 AGTCCTGACCAAGCTCCGACGGCGCTGGGCGGTAGCGGTTCGAGGCG 1009  
|||||  
34 InSerLeuThrLysLeuAlaAlaAlaTrpGlyGlySerGlySerGluAla 50  
1010 TAC 1012  
|||  
51 Tyr 51





151 GGCCTGCTGCGCGGAGTCCCTACCTGGCGCAGGTGGTGGTCTGTGACCCG 200  
|||||  
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292  
|||||  
201 CACCGCGTGTATCTCAGCTGATCGAAAGCCGCTTCCGCCCTCGGTGA 250  
|||||  
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309  
|||||  
251 TGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300  
|||||  
309 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325  
|||||  
301 GGTCCGGAGGAGTGGCGGAGGTTCGAATCCGGCGGCTCCACGAGCC 350  
|||  
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342  
|||||  
351 GGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 400  
|||||  
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAspG 359  
|||||  
401 AGGACGACTGGGACGAAGAGGAGGACTGG 429  
|||||  
359 LuAspAspTrpAspGluGluAspAspTrp 368  
|||||

seq\_name: /cgn2\_6/ptodata/2/paa/PCTUS\_COMB.pep:PCT-US99-03268-114

seq\_documentation\_block:  
; Sequence 114, Application PC/TUS9903268  
; GENERAL INFORMATION:  
; APPLICANT: Corixa Corporation  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY  
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS  
; NUMBER OF SEQUENCES: 355  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds, LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036-2811  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US99/03268  
; FILING DATE: 17-FEB-1999  
; CLASSIFICATION: 406  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 09/025,197  
; FILING DATE: 18-FEB-1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura, A.  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 9532-0013-228  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-493-4935  
; TELEFAX: 650-493-5556  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 114:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 368 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
PCT-US99-03268-114

alignment\_scores:  
Quality: 713.00 Length: 143  
Ratio: 5.057 Gaps: 0  
Percent Similarity: 98.601 Percent Identity: 96.503

alignment\_block:  
US-09-462-480-1 x PCT-US99-03268-114 ..  
Align seg 1/1 to: PCT-US99-03268-114 from: 1 to: 368  
1 CTGAGCAGGTGACGTCTGTTTTCAGCCAGGTGGCGGCACCCGGCGCGG 50  
|||||  
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGlyG 242  
|||||  
51 CAACCCACCCACGAGGAGCGCGCAGATGGCTGCTCGCGCACCATGTC 100  
|||||  
242 yAsnProAlaAspGluAlaAlaGlnMetGlyLeuLeuGlyThrSerP 259  
|||||  
101 CGCTGTGCAACATCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 150  
|||||  
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275  
|||||  
151 GGCCTGCTGCGCGGAGTCCCTACCTGGCGCAGGTGGTGGTCTGTGACCCG 200  
|||||  
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292  
|||||  
201 CACCGCGTGTATCTCAGCTGATCGAAAGCCGCTTCCGCCCTCGGTGA 250  
|||||  
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309  
|||||  
251 TGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300  
|||||  
309 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325  
|||||  
301 GGTCCGGAGGAGTGGCGGAGGTTCGAATCCGGCGGCTCCACGAGCC 350  
|||  
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342  
|||||  
351 GGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 400  
|||||  
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspAspG 359  
|||||  
401 AGGACGACTGGGACGAAGAGGAGGACTGG 429  
|||||  
359 LuAspAspTrpAspGluGluAspAspTrp 368  
|||||

seq\_name: /cgn2\_6/ptodata/2/paa/US086\_COMB.pep:US-08-658-800-109

seq\_documentation\_block:  
; Sequence 109, Application US/08658800  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF  
; NUMBER OF SEQUENCES: 123  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/658,800  
; FILING DATE: 05-JUN-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.417C3  
; TELECOMMUNICATION INFORMATION:

TUBERCULOSIS



seq\_name: /cgn2\_6/ptodata/2/paa/US086\_COMB.pep.US-08-680-573-109

seq\_documentation\_block:  
; Sequence 109, Application US/08680573  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Campos-Neto, Antonio  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF  
; NUMBER OF SEQUENCES: 128  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/680,573  
; FILING DATE: 12-JUL-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.417C4  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 109:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 368 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-680-573-109

alignment\_scores:  
Quality: 713.00 Length: 143  
Ratio: 5.057 Gaps: 0  
Percent Similarity: 98.601 Percent Identity: 96.503

alignment\_block:  
US-09-462-480-1 x US-08-680-573-109 ..

Align seg 1/1 to: US-08-680-573-109 from: 1 to: 368

1 CTCGACGAGTGAGTGGTGTTCAGCCAGGTGGCGGCGCACCGCGG 50  
|||||  
226 LeuGlnValThrSerLeuPheSerGlnValGlyThrGlyGly 242

51 CAACCCAGCCGACGAGGAAGCGCGGAGATGGCTCTCGGCACCAATC 100  
|||||  
242 YAsnProAlaAspGluGlnAlaGlnMetGlyLeuGlyThrSerP 259

101 CGCTGTCAACCATCCGCTGGCTGGTGGATCAGGCCCGCGCGCGCG 150  
|||||  
259 roLeuSerAsnHisProLeuAlaGlySerGlyProSerAlaGlyAla 275

151 GGCTGTGTCGGCGGAGTGGCTACCTGGCGGCGAGGTGGTGGTACCCG 200  
|||||  
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlySerLeuThrAr 292

201 CAGCCGCTGATGCTCAGCTGATCGAAAGCGGTTGGCCCTCGGTGA 250  
|||||  
292 gThrProLeuMetSerGlnLeuLeuGluLysProValAlaProSerValM 309

251 TGCCGGCGGCTGTTCGCGGATCGTGGGTGACGGGTGGCGGCGCTCCGGTG 300  
|||||  
309 etProAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325

301 GGTCCGGAGGATGGCCAGGTTCGCAATCCGCGGCTCCACCAGCCC 350  
|||||  
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342

351 GGGTCTGTGTCGCGCGGCGCACCGCTCGCGGAGGAGCGTGAAGAAGACGACG 400  
|||||  
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluAspAspG 359

401 AGGACGACTGGGACGAGAGGACGACTGG 429  
|||||  
359 luAspAspTrpAspGluGluAspTrp 368

seq\_name: /cgn2\_6/ptodata/2/paa/US086\_COMB.pep.US-08-680-574-114

seq\_documentation\_block:  
; Sequence 114, Application US/08680574  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Campos-Neto, Antonio  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY  
; NUMBER OF SEQUENCES: 133  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/680,574  
; FILING DATE: 12-JUL-1996  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.411C4  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 114:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 368 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-680-574-114

alignment\_scores:  
Quality: 713.00 Length: 143  
Ratio: 5.057 Gaps: 0  
Percent Similarity: 98.601 Percent Identity: 96.503

alignment\_block:  
US-09-462-480-1 x US-08-680-574-114 ..

Align seg 1/1 to: US-08-680-574-114 from: 1 to: 368

1 CTCGACGAGTGAGTGGTGTTCAGCCAGGTGGCGGCGCACCGCGGCGG 50  
|||||  
226 LeuGlnValThrSerLeuPheSerGlnValGlyThrGlyGly 242

```
51 CAACCCAGCGAGAGGAGCGCGCAGATGGGCTGCTCGGCACAGTC 100
|||||
242 YasnProAlaAspGluAlaAlaGlnMetGlyLeuLeuGlyThrSerP 259
|||||
101 CGCTGTCGAACCATCCGCTGGCTGGTGATCAGGCCCGCGGGCGG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlySerGlyProSerAlaGlyAla 275
|||||
151 GGCTGCTGCGCGGAGTCCCTACCTGCGCGCAGGTGGCTGTGACCG 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThr 292
|||||
201 CACGCCGCTGATGCTCAGCTGATCGAAGAGCGGTTGCCCGCTCGGTG 250
|||||
292 gThrProLeuMetSerGlnLeuLeuGlnProValAlaProSerValM 309
|||||
251 TGC CGCGGGCTGTTGCGGATCGCTGCTGAGCGGTGGCGCGCTCCGGTG 300
|||||
309 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
|||||
301 GGTTCGCGGAGCGATGGCGCAGGTTCGCAATCCGGCGGCTCCACGCGC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
|||||
351 GGCTGCTGCTGCGCGCACCGCTCGCGCAGGAGCGGTGAAGAACGACG 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspSpG 359
|||||
401 AGGACGACTGGGAGAGAGGAGGACTGG 429
|||||
359 luAspAspTrpAspGluGluAspTrp 368
|||||
```

seq\_name: /cgn2\_6/ptodata/2/paa/US087\_COMB.pap:US-08-729-622-109

```
seq_documentation_block:
; Sequence 109, Application US/08729622
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, David R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 132
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/729,622
; FILING DATE: 11-OCT-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C5
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 109:
; SEQUENCE CHARACTERISTICS:
```

```
; LENGTH: 368 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-729-622-109

alignment_scores:
Quality: 713.00 Length: 143
Ratio: 5.057 Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503

alignment_block:
US-09-462-480-1 x US-08-729-622-109 ..
Align seg 1/1 to: US-08-729-622-109 from: 1 to: 368

1 CTGACGAGGTGAGCTGCTGTTTCAGCCAGGTGGCGGCGCACCGCGCGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGly 242
|||||
51 CAACCCAGCGAGGAGGAGCGCGCAGATGGGCTGCTCGGCACAGTC 100
|||||
242 YasnProAlaAspGluAlaAlaGlnMetGlyLeuLeuGlyThrSerP 259
|||||
101 CGCTGTCGAACCATCCGCTGGCTGGTGATCAGGCCCGCGCGGCGCG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275
|||||
151 GGCTGCTGCTGCGCGCACCGCTACCTGCGCGCAGGTGGCTGTGACCG 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThr 292
|||||
201 CACGCCGCTGATGCTCAGCTGATCGAAGAGCGGTTGCCCGCTCGGTG 250
|||||
292 gThrProLeuMetSerGlnLeuLeuGlnProValAlaProSerValM 309
|||||
251 TGC CGCGGGCTGTTGCGGATCGCTGCTGAGCGGTGGCGCGCTCCGGTG 300
|||||
309 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
|||||
301 GGTTCGCGGAGCGATGGCGCAGGTTCGCAATCCGGCGGCTCCACGCGC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
|||||
351 GGCTGCTGCTGCGCGCACCGCTCGCGCAGGAGCGGTGAAGAACGACG 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspSpG 359
|||||
401 AGGACGACTGGGAGAGAGGAGGACTGG 429
|||||
359 luAspAspTrpAspGluGluAspTrp 368
|||||
```

seq\_name: /cgn2\_6/ptodata/2/paa/US087\_COMB.pap:US-08-730-510-114

```
seq_documentation_block:
; Sequence 114, Application US/08730510
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 137
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
```

seq\_name: /cgn2\_6/ptodata/2/paa/US089\_COMB.pep:US-08-942-341-109

seq\_documentation\_block:  
: Sequence 109, Application US/08942341  
: GENERAL INFORMATION:  
: APPLICANT: Reed, Steven G.  
: APPLICANT: Skeiky, Yasir A.W.  
: APPLICANT: Dillon, Davin C.  
: APPLICANT: Campos-Neto, Antonia  
: APPLICANT: Houghton, Raymond  
: APPLICANT: Vedvick, Thomas S.  
: APPLICANT: Twardzik, Daniel R.  
: APPLICANT: Loges, Michael J.  
: TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF  
: NUMBER OF SEQUENCES: 209  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: SEED and BERRY LLP  
: STREET: 6300 Columbia Center, 701 Fifth Avenue  
: CITY: Seattle  
: STATE: Washington  
: COUNTRY: USA  
: ZIP: 98104-7092  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: PatentIn Release #1.0, Version #1.30  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/730,510  
: FILING DATE: 27-AUG-1996  
: CLASSIFICATION: 424  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Maki, David J.  
: REGISTRATION NUMBER: 31,392  
: REFERENCE/DOCKET NUMBER: 210121.41105  
: TELEPHONE: (206) 622-4900  
: TELEFAX: (206) 682-6031  
: INFORMATION FOR SEQ ID NO: 114:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 368 amino acids  
: TYPE: amino acid  
: STRANDEDNESS: single  
: TOPOLOGY: linear  
: US-08-730-510-114

alignment\_scores:  
: Quality: 713.00 Length: 143  
: Ratio: 5.057 Gaps: 0  
Percent Similarity: 98.601 Percent Identity: 96.503

## alignment\_block:

US-09-462-480-1 x US-08-730-510-114

Align seg 1/1 to: US-08-730-510-114 from: 1 to: 368

1 CTGCAGCAGGTGACGTCTGTTTCAGCCAGGTGGCGGCACCGCGCGG 50  
|||||  
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGlyG 242  
51 CAACCCAGCGACGAGGAGCGCGGCGAGATGGCTCTCGGCACCGATC 100  
|||||  
242 YAsnProAlaAspGluGluAlaGlnMetGlyLeuLeuGlyThrSerP 259  
101 CGCTGTCTGAACCATCCGCTGCTGGTGGATCAGGCCCGCGCGCGCG 150  
|||||  
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275  
151 GGCCTGTCTGCGCGGAGTCGCTACCTGCGCGCAGGTGGTGGTGGACCG 200  
|||||  
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlySerLeuThrAr 292  
201 CACGCGCGTGATGCTCAGCTGATCGAAAGCGGTTGCCCGCTCGGTGA 250  
|||||  
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValm 309  
251 TGCCGCGCGCTGTTGCGCGATCGTCGCTGACGGTGGCGCGCTCCGGTG 300  
|||||  
309 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325  
301 GGTCCGCGGAGCGATGGCGAGGTTCGAATCCGCGCGCTCCACCGCCG 350  
|||||  
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlySerThrArgrPr 342  
351 GGGTCTGTGTCGCGCGCGACCGCTCGCGCAGGAGCGGTGAAGAGACGACG 400  
|||||  
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluAspaspG 359  
401 AGGACGACTGGGACGAGACGACGACTGG 429  
|||||  
359 luAspAspTrpAspGluGluAspAspTrp 368

TUBERCULOSIS



```
201 CACCCCGCTGATGCTCAGCTGATCGAAAGCCGGTTGCCCGCTCGGTGA 250
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValm 309
251 TCCCGCGCGCTGTTGCCGGATCGCTGCGTACGGGTGCGCGCTCCGGTG 300
309 etProAlaAlaAlaMetGlyGlnGlySerAlaThrGlyGlyAlaAlaProVal 325
301 GGTCCGGAGCGATGGCGGCGTCCAAATCCGGCGCTCCACCGACCC 350
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
351 GGGTCTGTCGCGCGCGACCGCTCGCGCAGGAGCGTGAAGAACGACG 400
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluAspAspG 359
401 AGGACGACTGGGAGAGAGACGACTGG 429
359 luAspAspTrpAspGluAspAspTrp 368
```

seq\_name: /cqn2\_6/ptodata/2/paa/US089\_COMB.pep:US-08-942-578-114

```
seq_documentation_block:
; Sequence 114, Application US/08942578
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 214
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/942,578
; FILING DATE: 01-OCT-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C7
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 114:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 368 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-942-578-114
```

alignment\_scores: Quality: 713.00 Length: 143  
Ratio: 5.057 Gaps: 0  
Percent Similarity: 98.601 Percent Identity: 96.503

```
alignment_block:
US-09-462-480-1 x US-08-942-578-114
Align seq 1/1 to: US-08-942-578-114 from: 1 to: 368
1 CTGCAGCAGGTGACGTGCTGTTTCAGCCAGTGGCGGCACCGCGCGCG 50
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGlyG 242
51 CACCCAGCGCAGCAGAGCCGCGCAGATGGCGCTGCTCGGCACCACTC 100
242 YasnProAlaAspGluAlaAlaGlnMetGlyLeuLeuGlyThrSerP 259
101 CGTGTCTCAACCATCCGCTGCTGGATCAGCCCGCCAGCGCGCGCG 150
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275
151 GGCTGTGTCGCGCGGAGTCTGCTACCTGGCGCAGGTGGGTGGTGA 200
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292
201 CACCCCGCTGATGCTCAGCTGATCGAAAGCCGGTTGCCCGCTCGGTGA 250
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValm 309
251 TCCCGCGCGCTGTTGCCGGATCGCTGCGTACGGGTGCGCGCTCCGGTG 300
309 etProAlaAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
301 GGTCCGGAGCGATGGCGCAGGTTCCAAATCCGGCGCTCCACCGACCC 350
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
351 GGTCTGTCGCGCGCGCGCTGCGCGCAGGAGCGTGAAGAACGACG 400
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluAspAspG 359
401 AGGACGACTGGGAGAGAGACGACTGG 429
359 luAspAspTrpAspGluAspAspTrp 368
```

seq\_name: /cqn2\_6/ptodata/2/paa/US090\_COMB.pep:US-09-024-753-109

```
seq_documentation_block:
; Sequence 109, Application US/09024753
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; NUMBER OF SEQUENCES: 236
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/024,753
; FILING DATE: 18-FEB-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
```

TUBERCULOSIS

```

; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C8
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 109:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 368 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-09-024-753-109

alignment_scores:
    Quality: 713.00      Length: 143
    Ratio: 5.057        Gaps: 0
    Percent Similarity: 98.601      Percent Identity: 96.503

alignment_block:
US-09-462-480-1 x US-09-024-753-109

Align seg 1/1 to: US-09-024-753-109 from: 1 to: 368

1 CTGCAGCAGGTGACGTCGTTGTTTCAGCCAGGTGGCGGCACCGCGCGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGlyG 242
51 CAACCCAGCCAGCAGGAGGAGCGGCGGCGGCGGCGGCGGCGGCGG 100
|||||
242 YAsnProAlaAspGluGluAlaGlnMetGlyLeuLeuGlyThrSerP 259
101 CGCTGTGCAACCATCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275
151 GGCTGTGCTGCGCGGAGTGCCTACCTACCTGCGGCGGCGGCGGCGG 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThr 292
201 CACGCCGCTGATGCTCAGCTGATCGAAGAGCGGCGGCGGCGGCGG 250
|||||
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerVal 309
251 TGCCGCGGCGGCTGTTGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300
|||||
309 etProAlaAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArg 400
351 GGGTCTGTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 429
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspG 359
401 AGGACGACTGGGCGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 429
359 luAspAspTrpAspGluGluAspAspTrp 368

seq_name: /cgn2_6/ptodata/2/paa/US090_COMB pep:US-09-025-197-114

seq_documentation_block:
; Sequence 114, Application US/09025197
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
```

```
351 GGGTGTGGTCCGCGGACCGCTCGCGCAGGAGCGTGAAGAACGACGACG 400
|||||
342 GGLyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspSPG 359
|||||
401 AGGACGACTGGGACGAGACGACTGG 429
|||||
359 luAspAspTrpAspGluAspAspTrp 368
```

seq\_name: /cgn2\_6/ptodata/2/paa/US090\_COMB.pep:US-09-072-596-109

```
seq_documentation_block:
; Sequence 109, Application US/09072596
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Twardzik, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 109:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 368 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-072-596-109
```

```
alignment_scores:
Quality: 713.00 Length: 143
Ratio: 5.057 Gaps: 0
Percent Similarity: 98.601 Percent Identity: 96.503

alignment_block:
US-09-462-480-1 x US-09-072-596-109 ..
Align seg 1/1 to: US-09-072-596-109 from: 1 to: 368

1 CTGCAGCAGTGACGCTGTTTTCAGCCAGGTGGCGGACCGCGGCGG 50
|||||
226 LeuGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGly 242
|||||
51 CAACCCAGCCGACGAGGAGCGCGCAGATGGCTGCTGCACCACTC 100
|||||
```

```
242 yAsnProAlaAspGluGluAlaAlaGlnMetGlyLeuLeuGlyThrSerP 259
|||||
101 CGCTGTCTGAACCATCCGCTGCTGGTGTGATCAGCCAGCCCGCGGCGCG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlySerGlyProSerAlaGlyAla 275
|||||
151 GGCCTGTGTCGCGCGGAGTCTACCTGGGCGAGTGGGTGCTGTTGACCCG 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292
|||||
201 CACGCCCTGATGTCTCAGCTGATCGAAAGCCGCTGTCGCCCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM 309
|||||
251 TGC CGCGGCTGTTTGGCGGATCGCTGCGTACGCGGTGGCGGCTCCCGTG 300
|||||
309 etProAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
|||||
301 GGTCCGGGAGCGATGGCCAGGGTTCGCAATCCGGCGGCTCCACCGCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
|||||
351 GGTCTGTGTCGCGCGGCGGACCGCTCGCGCAGGAGCGTGAAGAACGACG 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluAspSPG 359
|||||
401 AGGACGACTGGGACGAGACGACTGG 429
|||||
359 luAspAspTrpAspGluAspAspTrp 368
```

```
seq_name: /cgn2_6/ptodata/2/paa/US090_COMB.pep:US-09-072-967-114

seq_documentation_block:
; Sequence 114, Application US/09072967
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Twardzik, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 114:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 368 amino acids
```

TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-072-967-114

alignment\_scores:  
Quality: 713.00 Length: 143  
Ratio: 5.057 Gaps: 0  
Percent Similarity: 98.601 Percent Identity: 96.503

alignment\_block:  
US-09-462-480-1 x US-09-072-967-114

Align seg 1/1 to: US-09-072-967-114 from: 1 to: 368.

```
1 CTGCAGCAGGTGACGTCGTTGTTTTCAGCCAGGTGGCGGCACCGCGCGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGlyG 242
51 CAACCCAGCCGACGAGGAAGCGCGCAGATGGGCTGCTCGGCACCGAGTC 100
|||||
242 yAsnProAlaAspGluGluAlaGlnMetGlyLeuLeuGlyThrSerP 259
|||||
101 CGCTGTGCAACCATCCGCTGGCTGGTGATCAGGCCCGCCGCGCGCG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275
151 GGCCTGTGTCGCGGAGTCGCTACCTGGCGCAGGTGGGTGCTTGACCCG 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThr 292
|||||
201 CACGCGCTGATGTCAGCTGATCGAAAGCCGGTTGCCCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuLeuGluLysProValAlaProSerVal 309
|||||
251 TGCCGCGCGCTGTTCCCGAGTCGTCGGTGACGGTGGCGCGCGCTCGGT 300
|||||
309 etProAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
|||||
301 GGTCCGGGAGGATGGCCAGGTTGCAATCCGCGCGCTCCACCGCC 350
|||||
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArg 342
|||||
351 GGGTCTGTGTCGCGCGGACCGCTCGCGCAGGAGCGGTGAAGAAGCAGC 400
|||||
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluAspAspG 359
|||||
401 AGGACGACTGGGACGAGAGGACGACTGG 429
|||||
359 luAspAspTrpAspGluGluAspTrp 368
```

seq\_name: /cqn2\_5/ptodata/2/paa/US097\_COMB.pep:US-09-724-685-114

seq\_documentation\_block:

Sequence 114, Application US/09724685

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.

Skeiky, Yasir  
Dillon, Davin C.

Campos-Neto, Antonio

TITLE OF INVENTION: Compounds and Methods for

Immunotherapy and Diagnosis of Tuberculosis

NUMBER OF SEQUENCES: 155

CORRESPONDENCE ADDRESS:

ADDRESSEE: Townsend and Crew LLP

STREET: Two Embarcadero Center, Eighth Floor

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-3834

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/724,685  
FILING DATE: 28-Nov-2000  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/523,436  
FILING DATE: 01-SEP-1995  
APPLICATION NUMBER: US 08/533,634  
FILING DATE: 22-SEP-1995  
APPLICATION NUMBER: US 08/620,874  
FILING DATE: 22-MAR-1996  
APPLICATION NUMBER: US 08/659,683  
FILING DATE: 05-JUN-1996  
APPLICATION NUMBER: US 08/680,574  
FILING DATE: 12-JUL-1996  
APPLICATION NUMBER: WO PCT/US96/14674  
FILING DATE: 30-AUG-1996  
APPLICATION NUMBER: US 08/730,511  
FILING DATE: 11-OCT-1996  
APPLICATION NUMBER: US 08/818,112  
FILING DATE: 13-MAR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Bastian, Kevin L.  
REGISTRATION NUMBER: 34,774  
REFERENCE/DOCKET NUMBER: 014058-008561US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 114:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 368 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 114:  
US-09-724-685-114

alignment\_scores:

Quality: 713.00 Length: 143  
Ratio: 5.057 Gaps: 0

Percent Similarity: 98.601 Percent Identity: 96.503

alignment\_block:

US-09-462-480-1 x US-09-724-685-114

Align seg 1/1 to: US-09-724-685-114 from: 1 to: 368

```
1 CTGCAGCAGGTGACGTCGTTGTTTTCAGCCAGGTGGCGGCACCGCGCGG 50
|||||
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGlyG 242
51 CAACCCAGCCGACGAGGAAGCGCGCAGATGGGCTGCTCGGCACCGAGTC 100
|||||
242 yAsnProAlaAspGluGluAlaGlnMetGlyLeuLeuGlyThrSerP 259
|||||
101 CGCTGTGCAACCATCCGCTGGCTGGTGATCAGGCCCGCCGCGCGCG 150
|||||
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275
151 GGCCTGTGTCGCGGAGTCGCTACCTGGCGCAGGTGGGTGCTTGACCCG 200
|||||
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThr 292
|||||
201 CACGCGCTGATGTCAGCTGATCGAAAGCCGGTTGCCCTCGGTGA 250
|||||
292 gThrProLeuMetSerGlnLeuLeuGluLysProValAlaProSerVal 309
|||||
251 TGCCGCGCGCTGTTCCCGAGTCGTCGGTGACGGTGGCGCGCGCTCGGT 300
|||||
```



**THIS PAGE BLANK (USPTO)**

OM of: US-09-462-480-1 to: Pending\_Patents\_AA\_New:\* out\_format : pfs  
Date: Jul 22, 2002 1:37 AM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:  
-MODEL=frame+\_n2p.model -DEV=xlh  
-O=/cgn2\_1/uspt0.spool/US09462480/runat\_18072002\_164419\_19498/app\_query.fasta\_1.2850  
-DB=Pending\_Patents\_AA\_New -QFMT=fastn -SUFFIX=rapn  
-GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000  
-LOOPEXT=0.000 -QGAPOP=4.500 -QGAPOP=10.000  
-XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000  
-YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1  
-MATRIX=blosom62 -TRANS=human40.cdi -LIST=45 -DICALIGN=200  
-THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=LOCAL  
-OUTFMT=pfs -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09462480 -CGNJ\_1\_93 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

Search information block:  
Query: US-09-462-480-1  
Query length: 1277  
Database: Pending\_Patents\_AA\_New:\*  
Database sequences: 378952  
Database length: 124292526  
Search time (sec): 104.950000

score_list:	Sequence	Strd Orig	zscore	Escore	Len	Documentation
/cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-116-492A-5	+	492.00	493.20	2.9e-19	10	
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-140-045-5	+	492.00	493.20	2.9e-19	100	
/cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-10-080-170B-639	+	486.00	487.37	6.0e-19	49	
/cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-116-492A-6	+	247.00	260.53	5.3e-06	49	
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-140-045-6	+	247.00	260.53	5.3e-06	49	
/cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-935-625-17077	-	210.50	202.33	0.0006	49	
/cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-935-625-17076	-	210.50	202.13	0.0006	49	
/cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-935-625-17480	-	210.50	202.02	0.0006	49	
/cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-935-625-17479	-	210.50	201.84	0.0006	49	
/cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-935-625-17075	-	210.50	201.73	0.0006	49	
/cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-935-625-17478	-	210.50	201.62	0.0006	49	
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-09-116-492A-7	+	205.00	220.89	0.0010	42	
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-140-045-7	+	205.00	220.89	0.0010	42	
/cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep:US-60-389-987-2231	+	203.50	187.43	0.0014	42	
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-080-170B-13	+	198.00	198.25	0.0025	3	
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-155-881-10704	-	191.50	190.60	0.0058	3	
/cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-116-492A-28	+	187.00	196.41	0.0097	10	
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-140-045-28	+	187.00	196.41	0.0097	10	
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-080-170B-12	+	187.00	196.41	0.0097	10	
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-080-170B-11	+	186.00	195.85	0.0110	10	
/cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep:PCT-US02-13142-3473	-	181.50	170.81	0.0208	9	
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-128-714-3473	-	181.50	170.81	0.0208	9	
/cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep:PCT-US02-13142-8473	-	175.50	164.87	0.0440	9	
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-128-714-8473	-	175.50	164.87	0.0440	9	
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-155-881-26077	+	172.50	171.65	0.0517	9	
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-155-881-26078	+	172.50	171.65	0.0517	9	
/cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-881-26387	+	171.00	166.24	0.0755	9	
/cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-667-170A-187	-	167.00	162.64	0.1242	5	
/cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep:US-60-365-264-664	-	166.50	162.80	0.1319	5	
/cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep:US-60-360-039-20627	-	165.00	158.96	0.1605	5	
/cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep:US-60-389-987-2448	-	165.00	156.09	0.1622	5	
/cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-582-687A-60	-	164.50	154.68	0.1732	1	
/cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-935-625-17480	+	164.00	156.77	0.1826	1	
/cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-935-625-17479	+	164.00	156.59	0.1828	1	
/cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-935-625-17478	+	164.00	156.37	0.1829	1	
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-155-881-11792	-	163.50	163.09	0.1896	1	
/cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-121-988-47	-	163.50	156.60	0.1942	82	
/cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep:US-60-389-987-2272	-	162.00	152.22	0.2367	82	
/cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep:US-60-389-987-2617	+	161.00	158.44	0.2611	82	
/cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep:US-10-155-881-10213	+	161.00	156.95	0.2626	82	
/cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep:US-60-360-039-1813	-	161.00	154.26	0.2652	82	

/cgn2\_6/ptodata/2/paa/US60\_NEW\_COMB.pep:US-60-389-987-2231 - 159.50 144.62 0.3295  
/cgn2\_6/ptodata/2/paa/US09\_NEW\_COMB.pep:US-09-994-404-252 - 159.00 141.73 0.3538  
/cgn2\_6/ptodata/2/paa/PCT\_NEW\_COMB.pep:PCT-US02-19669-70 + 158.50 146.97 0.3688  
/cgn2\_6/ptodata/2/paa/US09\_NEW\_COMB.pep:US-09-935-625-9406 - 158.50 146.48 0.3694  
/cgn2\_6/ptodata/2/paa/US09\_NEW\_COMB.pep:US-09-935-625-26246 - 158.50 146.48 0.3694  
seq\_name: /cgn2\_6/ptodata/2/paa/US09\_NEW\_COMB.pep:US-09-116-492A-5  
seq\_documentation\_block:  
; Sequence 5, Application US/09116492A  
; GENERAL INFORMATION:  
; APPLICANT: GICQUEL, BRIGITTE  
; APPLICANT: BERTHET, FRANCOIS-XAVIER  
; APPLICANT: ANDERSEN, PETER  
; APPLICANT: RASMUSSEN, PETER B  
; TITLE OF INVENTION: POLYNUCLEOTIDE FUNCTIONALLY CODING FOR THE LHP PROTEIN FROM M  
; TITLE OF INVENTION: TUBERCULOSIS ITS BIOLOGICALLY ACTIVE DERIVATIVE FRAGMENTS,  
; TITLE OF INVENTION: USING THE SAME  
; FILE REFERENCE: 0660-0137-27X  
; CURRENT APPLICATION NUMBER: US/09/116,492A  
; CURRENT FILING DATE: 1998-07-16  
; PRIOR APPLICATION NUMBER: 60/252,631  
; PRIOR FILING DATE: 1997-07-16  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 5  
; LENGTH: 100  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis  
US-09-116-492A-5

alignment_scores:				
Quality:	492.00	Length:	100	
Ratio:	4.920	Gaps:	0	
Percent Similarity:	100.000	Percent Identity:	100.000	
alignment_block:				
US-09-462-480-1 x US-09-116-492A-5	..			
Align seg 1/1 to: US-09-116-492A-5	from: 1	to: 100		
525 ATGCAGAGATGAACACGATGCCGCTACCTCGGCGAGGAGGATAA	574			
1 MetAlaGluMetLysThrAspAlaAlaThrLeuGlyGlnGluAlaGlyAs	17			
575 TTTCGAGCGGATCTCCGCGCACCTGAAACCCAGATCGACACAGGTGAGT	624			
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGluS	34			
625 CGACGGCAGGTTCTGTTTCAGGCGCGAGTGGCGGCGGCGGCGGCC	674			
34 erThrAlaGlySerLeuGlnGlyGlnIleArgGlyAlaAlaGlyThrAla	50			
675 GCCAGGCGCGGTGGTGGCGTTCACCAAGACGACCAATAAGCAGAAGCA	724			
51 AlaGlnAlaAlaValAlaArgPheGlnGluAlaAlaAsnLysGlnLysG	67			
725 GGAACTCGACGAGATCTCGACGAATATTCGTCAGCGCGCGTCCAATACT	774			
67 nGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyrS	84			
775 CGAGGCGCGAGGAGGACGACGAGCGGCTGCTCTGCAATGGGCTTC	824			
84 erArgAlaAspGluGlnGlnGlnAlaLeuSerGlnMetGlyPhe	100			
seq_name: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-140-045-5				
seq_documentation_block:				
; Sequence 5, Application US/10140045				
; GENERAL INFORMATION:				
; APPLICANT: GICQUEL, BRIGITTE				
; APPLICANT: BERTHET, FRANCOIS-XAVIER				

```
; APPLICANT: ANDERSEN, PETER
; APPLICANT: RASMUSSEN, PETER B
; TITLE OF INVENTION: POLYNUCLEOTIDE FUNCTIONALLY CODING FOR THE LHP PROTEIN FROM MYCOBACTERIUM TUBERCULOSIS, ITS BIOLOGICALLY ACTIVE DERIVATIVE FRAGMENTS, AS WELL AS A TOOL FOR THE DIAGNOSIS, PROPHYLAXIS OR TREATMENT OF MYCOBACTERIOSES
; TITLE OF INVENTION: USING THE SAME
; FILE REFERENCE: 0660-0137-27X
; CURRENT APPLICATION NUMBER: US/10/140,045
; PRIOR FILING DATE: 2002-05-08
; PRIOR APPLICATION NUMBER: US/09/116,492A
; PRIOR FILING DATE: 1998-07-16
; PRIOR APPLICATION NUMBER: 60/252,631
; PRIOR FILING DATE: 1997-07-16
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 100
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-140-045-5

alignment_scores:
  Quality: 492.00      Length: 100
  Ratio: 4.920        Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-462-480-1 x US-10-140-045-5 ..
Align seg 1/1 to: US-10-140-045-5 from: 1 to: 100

525 ATGCAGAGATGAAGACCGATGCGCTACCTCGGCGAGGAGCGGTAA 574
|||||
1 MetAlaGluMetLysThrAspAlaAlaThrLeuGlnGluAlaGlyAs 17

575 TTTCAGCGGATCTCCGGCGACCTGAAACCCAGATCGACCGAGTGGAGT 624
|||||
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGluS 34

625 CGACGGCGGTTGTTGTCAGGCGGAGTGGCGGCGGCGGCGGCGGCC 674
|||||
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50

675 GCCAGCGCGGTTGTCGCTTCCAAAGACGAGCAATAAGCAGAAACA 724
|||||
51 AlaGlnAlaAlaValAlaArgPheGlnGluAlaAlaAsnLysGlnLysG 67

725 GGAACCTCGACGATCTCGACGATATTCGTCAGGCGGCGCTCCCAATACT 774
|||||
67 nGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyrS 84

775 CGAGGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGTTC 824
|||||
84 erArgAlaAspGluGlnGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 100

seq_name: /cn2_6/ptodata/2/paa/US10_NEW_COMB.pep:US-10-080-170B-639

seq_documentation_block:
; Sequence 639, Application US/10080170B
; GENERAL INFORMATION:
; APPLICANT: COLE, S.T.
; TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR TREATMENT OF MYCOBACTERIOSES
; FILE REFERENCE: 03495.0218
; CURRENT APPLICATION NUMBER: US/10/080,170B
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: 60/270,123
; PRIOR FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 652
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 639
; LENGTH: 100
```

```
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-080-170B-639

alignment_scores:
  Quality: 486.00      Length: 100
  Ratio: 4.860        Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 99.000

alignment_block:
US-09-462-480-1 x US-10-080-170B-639 ..
Align seg 1/1 to: US-10-080-170B-639 from: 1 to: 100

525 ATGCAGAGATGAAGACCGATGCGCTACCTCGGCGAGGAGCGGTAA 574
|||||
1 MetAlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAs 17

575 TTTCAGCGGATCTCCGGCGACCTGAAACCCAGATCGACCGAGTGGAGT 624
|||||
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGluS 34

625 CGACGGCGGTTGTTGTCAGGCGGAGTGGCGGCGGCGGCGGCGGCC 674
|||||
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50

675 GCCAGCGCGGTTGTCGCTTCCAAAGACGAGCAATAAGCAGAAACA 724
|||||
51 AlaGlnAlaAlaValAlaArgPheGlnGluAlaAlaAsnLysGlnLysG 67

725 GGAACCTCGACGATCTCGACGATATTCGTCAGGCGGCGCTCCCAATACT 774
|||||
67 nGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTyrS 84

775 CGAGGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGTTC 824
|||||
84 erArgAlaAspGluGlnGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 100

seq_name: /cn2_6/ptodata/2/paa/US09_NEW_COMB.pep:US-09-116-492A-6

seq_documentation_block:
; Sequence 6, Application US/09116492A
; GENERAL INFORMATION:
; APPLICANT: GICQUEL, BRIGITTE
; APPLICANT: BERTHET, FRANCOIS-XAVIER
; APPLICANT: ANDERSEN, PETER
; APPLICANT: RASMUSSEN, PETER B
; TITLE OF INVENTION: POLYNUCLEOTIDE FUNCTIONALLY CODING FOR THE LHP PROTEIN FROM MYCOBACTERIUM TUBERCULOSIS, ITS BIOLOGICALLY ACTIVE DERIVATIVE FRAGMENTS, A TOOL FOR THE DIAGNOSIS, PROPHYLAXIS OR TREATMENT OF MYCOBACTERIOSES
; FILE REFERENCE: 0660-0137-27X
; CURRENT APPLICATION NUMBER: US/09/116,492A
; CURRENT FILING DATE: 1998-07-16
; PRIOR APPLICATION NUMBER: 60/252,631
; PRIOR FILING DATE: 1997-07-16
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 49
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-116-492A-6

alignment_scores:
  Quality: 247.00      Length: 49
  Ratio: 5.041        Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-462-480-1 x US-09-116-492A-6 ..
```



Align seg 1/1 to: US-09-116-492A-6 from: 1 to: 49

```
525 ATGCAGAGATGAACACCGATGCCGCTACCTCGGCGAGGAGGAGGTAA 574
|||||
1 MetAlaGluMetLysThrAspAlaAlaThrLeuGlnGluAlaGlyAs 17
|||||
575 TTTCGACGGATCTCCGGCGACCTGAAACCCAGATCGACCGAGTGGAGT 624
|||||
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGluS 34
|||||
625 CGACGGCAGGTCTTCGACGGCGCAGTGGCGCGCGCGGGGAGC 671
|||||
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThr 49
```

seq\_name: /cgn2\_6/ptodata/2/paa/US10\_NEW\_COMB.pap:US-10-140-045-6

```
seq_documentation_block:
; Sequence 6, Application US/10140045
; GENERAL INFORMATION:
; APPLICANT: GICQUEL, BRIGITTE
; APPLICANT: BERTHET, FRANCOIS-XAVIER
; APPLICANT: ANDERSEN, PETER
; APPLICANT: RASMUSSEN, PETER B
; TITLE OF INVENTION: POLYNUCLEOTIDE FUNCTIONALLY CODING FOR THE LHP PROTEIN FROM MYCOBACTERIUM TUBERCULOSIS, ITS BIOLOGICALLY ACTIVE DERIVATIVE FRAGMENTS, AS WELL AS METHODS THEREOF CAPABLE OF MODULATING VARIOUS RESPONSES
; TITLE OF INVENTION: USING THE SAME
; FILE REFERENCE: 0660-0137-27X
; CURRENT APPLICATION NUMBER: US/10/140,045
; PRIOR APPLICATION NUMBER: 2002-05-08
; PRIOR FILING DATE: 1998-07-16
; PRIOR APPLICATION NUMBER: 60/252,631
; PRIOR FILING DATE: 1997-07-16
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-140-045-6
```

alignment\_scores:  
Quality: 247.00 Length: 49  
Ratio: 5.041 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-462-480-1 x US-10-140-045-6 ..

Align seg 1/1 to: US-10-140-045-6 from: 1 to: 49

```
525 ATGCAGAGATGAACACCGATGCCGCTACCTCGGCGAGGAGGAGGTAA 574
|||||
1 MetAlaGluMetLysThrAspAlaAlaThrLeuGlnGluAlaGlyAs 17
|||||
575 TTTCGACGGATCTCCGGCGACCTGAAACCCAGATCGACCGAGTGGAGT 624
|||||
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGluS 34
|||||
625 CGACGGCAGGTCTTCGACGGCGCAGTGGCGCGCGGGGAGC 671
|||||
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThr 49
```

seq\_name: /cgn2\_6/ptodata/2/paa/US09\_NEW\_COMB.pap:US-09-935-625-17077

```
seq_documentation_block:
; Sequence 17077, Application US/09935625
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: POLYNUCLEOTIDES, POLYPEPTIDES, CELLS, AND METHODS THEREOF CAPABLE OF MODULATING VARIOUS RESPONSES
; FILE REFERENCE: 2750-1481P
```

```
; CURRENT APPLICATION NUMBER: US/09/935,625
; CURRENT FILING DATE: 2001-08-24
; NUMBER OF SEQ ID NOS: 33136
; SEQ ID NO 17077
; LENGTH: 827
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: peptide
; LOCATION: 1...827
; OTHER INFORMATION: Ceres Seq. ID no. 2708208
US-09-935-625-17077
```

alignment\_scores:  
Quality: 210.50 Length: 380  
Ratio: 1.126 Gaps: 23  
Percent Similarity: 49.211 Percent Identity: 31.316

alignment\_block:  
US-09-462-480-1/rev x US-09-935-625-17077 ..

Align seg 1/1 to: US-09-935-625-17077 from: 1 to: 827

```
1253 TCTCAGAGTGGCTCAACGCTATAACACGAGAAAGGCGAGACGCGG 1204
|||||
234 SerArgSerProIleArgArg.....HisArgArgProThrHisGluL 248
|||||
1203 AAGTCGAACCTCGCCGATCCCGTGTTCGCTATTCTACGGAACCTCGC 1154
|||||
248 yArgArgGlnSerProAlaPro.....SerArgArgArgA 260
|||||
1153 GTTCGCCCTATCGCAACATCCCGATGAGTTCCTTCGGTCCGACCATG 1104
|||||
260 rgSerProSerProProAlaArgArg.....ArgSerProSerPro 274
|||||
1103 CCTGACCGGCTTCGCTATCGTCGCGCGAGTTCGAGCGCGTGTTC 1054
|||||
275 ProAlaArgArgArgArgSerProSerProProAlaArgArgHisArg 291
|||||
1053 AGTCGGTAGCGGTGGCGTCCCATTTTGTGGACACCTCGTAGCGCTC 1004
|||||
291 rProThrProProAlaArgGln.....ArgArgSerProSerProA 306
|||||
1003 CGAACCGCTACCGC...CCGAGCGCGCTCGAGCTTGGTCAGGAGTCT 957
|||||
306 laArgArgHisArgSerProProAlaArgArgArgArgSerProSer 322
|||||
956 TCCCTCGTCAAGGAGGGAATGAATGGACGTGACATTTCCCTGGATT 907
|||||
323 ProProAlaArgArgArgSerProSerProProAlaArgArgArgAr 339
|||||
906 CTTCCGCGGCGCTCGATACCGCGAAATTCACCTGCTGCTGCTCATGTT 857
|||||
339 gSerProSerProLeuYrArgArgAsnArgSerPro..SerProLeu 355
|||||
856 TTTCGCTCGTTTCTTCGTATTAGCGGTCCAGAACCCATTTTCGAGGA 807
|||||
355 rArgArgAsn.....ArgSerArgSerProLeuAlaLysA 367
|||||
806 CAGCGCTGCTGCTGCTCTCT...CGTCGCGCTCGAGTATTGGAGCG 760
|||||
367 rgGlyArgSerAspSerProGlyArgSerProSerProValAlaArg 383
|||||
759 CCTGACGAATATTGCTCGAGA.....TCTGCTCGAGT 728
|||||
384 ArgAspProThrGlyAlaArgLeuProSerProSerIleGluGlnArg 400
|||||
727 TCCTGCTCTCTGCTTATTGGCTGCTTCTTGAAGCGCACCG...CG 681
|||||
400 uProSerProProValAlaGlnArgLeuProSerProProProArg 417
|||||
680 CTGGCGCGCGCTCCCGCGCGCGCGCTGCGCTGCAACGAACCTG 631
```

```

417 laglyLeuProSerProPro.....ProAlaGlnArgLeu 428
630 CCGTCGACTCCACCTGGTCGATCTGGGTTTTCAGGTCGCCGAGG..... 586
429 ProSerProProProArgAlaGlyLeuProSerProMetArgIleG1 445
595 .....TCCGCTCGAATACCTG..... 568
445 yGlySerHisAlaAlaAsnHisLeuSerProSerProSerSerLeuS 462
567 ..CTCCGTCGCCGAGGTAGCGCATCGCTTTCATCTCTCCCATGCTGG 520
462 erProProGlyArg.....LysLysValLeuProSerProProValArg 476
519 ACTACTTCTCTCTTACCTTCTCCGCGCAAAATGTTGGCAAGTCTCCGG 470
477 ArgArgArgSerLeuThrProAspGluGluValSerLeuSerGln... 492
469 CCGGGTGGCGGGAAGTCTGTTGCAATTAC.....GGGAGC. 433
493 ...GlyGlyArgHisThrProSerHisIleLysGlnAspGlySerM 508
432 ..TCACCACTCGTCTCTCTTCGTCGAGTCGCTGCTCTTCTCAGC 385
508 etSerProVal..ArgGlyArgGlyLysSerProSerSerArg.HisG 524
384 CTCTGCGCGAGCGTGGCGCGGACGACCGG.....GGCTGG 344
524 InLysAlaArgSerProValArgArgSerProThrProValAsnArg 540
343 TGGAGCCCGCGGATGGAGACCTGGGCCATCGCTCCGCGGACCCACCGGA 294
541 ArgSerArgArgSerSerSerAlaSerArgSer...ProAspArgArgAr 556
293 GCGGCGCACCCG.....TCACGACGATCGGCAACAGCCGC 256
556 gArgArgSerProSerSerArgSerProSerArgSerArgSerProp 573
255 CGGCA.....TCACCGAGGGGGCAA 236
573 roValLeuHisArgSerProSerProArgGlyArg 584
```

seq\_name: /cgn2\_6/ptodata/2/paa/us09\_NEW\_COMB.pep.us-09-935-625-17076

```

seq_documentation_block:
; Sequence 17076, Application US/0935625
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: POLYNUCLEOTIDES, POLYPEPTIDES, CELLS, AND METHODS THEREOF CAPABLE
; FILE OF INVENTION: MODULATING VARIOUS RESPONSES
; FILE REFERENCE: 2750-1481P
; CURRENT APPLICATION NUMBER: US/09/935,625
; CURRENT FILING DATE: 2001-08-24
; NUMBER OF SEQ ID NOS: 33136
; SEQ ID NO 17076
; LENGTH: 847
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: peptide
; LOCATION: 1..847
; OTHER INFORMATION: Ceres Seq. ID no. 2708207
US-09-935-625-17076
```

alignment\_scores:  
Quality: 210.50 Length: 380  
Ratio: 1.126 Gaps: 23  
Percent Similarity: 49.211 Percent Identity: 31.316

alignment\_block:  
US-09-462-480-1/rev x US-09-935-625-17076 ..

```

Align seg 1/1 to: US-09-935-625-17076 from: 1 to: 847
1253 TCTCAGACTGGCTCAAAAGTATAACACGAGAACGGCGAGACCGG 1204
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
254 SerArgSerProIleArgArg.....HisArgArgProThrHisGluG1 268
1203 AAGTCGAACTCGCCCGCTCCGCTGTTTTCGCTATTCTACGCGAAGTCCGC 1154
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
268 yArgArgGlnSerProAlaPro.....SerArgArgArgA 280
1153 GTTGGCCCTATCGGAACATCCAGTCACGTTGCCCTTCGTCGAGCCATTG 1104
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
280 rgSerProSerProProAlaArgArg.....ArgSerProSerPro 294
1103 CCTGACCGCTTCGCTGATGTCGCCGCGCAGGTTCTCGAGCGCTGTTTC 1054
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
295 ProAlaArgArgArgArgSerProSerProProAlaArgArgHisArgSe 311
1053 AGCTCGGTAGCGGTGCGCTCCCATTTTTCGTGGACACCTGTCAGCCTC 1004
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
311 rProThrProProAlaArgGln.....ArgArgSerProSerProProA 326
1003 CGAACCGCTACCGC...CCCAGGCGCTGCGAGCTTGGTCAGGGACTGCT 957
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
326 laArgArgHisArgSerProProProAlaArgArgArgArgSerProSer 342
956 TCCCTCTCTCAAGGAGGAATGAGTGGACGTGACATTTCCCTGGATTGCG 907
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
343 ProProAlaArgArgArgArgSerProSerProProAlaArgArgArgAr 359
906 CTTCGCGCGGCTCGATACCGCGGAAATTCACCTGCTCTGCTCATGTT 857
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
359 gSerProSerProLeuTyArgArgAsnArgSerPro..SerProLeuTy 375
856 TTTGCTCGCTTCTTTCGTTATTAGCGGCTCAGAACCCCATTTGCGAGGA 807
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
375 rArgArgAsn.....ArgSerArgSerProLeuAlaLysA 387
806 CAGCGCTCTGCTGCTCTCT...CGTCGCGCTCGAGTATTGGAGCGCGG 760
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
387 rgGlyArgSerAspSerProGlyArgSerProSerProValAlaArgLeu 403
759 CCGTACGAATATTCGTCGAGA.....TCTCGTCGAGT 728
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
404 ArgAspProThrGlyAlaArgLeuProSerProSerIleGluGlnArgLe 420
727 TCCTGCTTCTCTTATTTGGCTGCTTCTTGGAGCGCACCCG...CGCG 681
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
420 uProSerProProValAlaGlnArgLeuProSerProProProArgArgA 437
680 CTGGCGGCGCTCCCGCGCGCGCGCTGCGCTGCGCAACGAACTCG 631
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
437 laglyLeuProSerProPro.....ProAlaGlnArgLeu 448
630 CCGTCGACTCCACCTGGTCGATCTGGGTTTTCAGGTCGCCGAGGAG... 586
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
449 ProSerProProArgArgAlaGlyLeuProSerProMetArgIleG1 465
585 .....TCCGCTCGAATACCTG..... 568
465 yGlySerHisAlaAlaAsnHisLeuGluSerProSerProSerSerLeuS 482
567 ..CCTCCTGCGCGAGGTAGCGCATCGGCTTCTTCATCTCTCCCATGCTGG 520
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
482 erProProGlyArg.....LysLysValLeuProSerProProValArg 496
519 ACTACTTCTCTCTTACCTTCTCCGCGCAAAATGTTGGCAAGTCTTCCGG 470
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
497 ArgArgArgSerLeuThrProAspGluGluArgValSerLeuSerGln.. 512
469 CCGGGTGGCGGGAAGTCTGTTGTCATTAC.....GGGAGC. 433
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
```

```
513 ....GlyGlyArgHisThrSerProSerHisIleLysGlnAspGlySerM 528
432 ..TCACAGTCGCTCTTCGTCCTCCAGTCGTCCTCGCTCTTCTTCACG 385
||||| ||| : : : : : ||| ||| ||| ||| ||| |||
528 etSerProVal.ArgGlyArgGlyLysSerSerProSerSerArg.HisG 544
||||| ||| : : : : : ||| ||| ||| ||| ||| |||
384 CTCCTCGCGAGCGGTGCCGCGCAGCACACCGG.....GGCTGG 344
||||| ||| : : : : : ||| ||| ||| ||| ||| |||
544 lnLysAlaArgSerProValArgArgSerProThrProValAsnArg 560
||||| ||| : : : : : ||| ||| ||| ||| ||| |||
343 TGGAGCGCGCGGATGGAACTCGTCCGATCGTCCCGGACCGACCGGA 294
||||| ||| : : : : : ||| ||| ||| ||| ||| |||
561 ArgSerArgArgSerSerSerAlaSerArgSer....ProAspArgArg 576
||||| ||| : : : : : ||| ||| ||| ||| ||| |||
293 GCGGCGCCACCGG.....TCACGAGCATCCGCGCAACAGCGCG 256
||||| ||| : : : : : ||| ||| ||| ||| ||| |||
576 gArgArgSerProSerSerSerArgSerProSerArgSerArgSerProp 593
||||| ||| : : : : : ||| ||| ||| ||| ||| |||
255 CGGCA.....TCACCGAGGGGGCAA 236
||||| ||| : : : : : ||| ||| ||| ||| ||| |||
593 roValLeuHisArgSerProSerProArgGlyArg 604
```

seq\_name: /cgn2\_6/ptodata/2/paa/US09\_NEW\_COMB pep: US-09-935-625-17480

```
seq_documentation_block:
; Sequence 17480, Application US/09935625
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: POLYNUCLEOTIDES, POLYPEPTIDES, CELLS, AND METHODS THEREOF CAPABLE
; TITLE OF INVENTION: MODULATING VARIOUS RESPONSES
```

```
; FILE REFERENCE: 2750-1481P
; CURRENT APPLICATION NUMBER: US/09/935,625
; CURRENT FILING DATE: 2001-08-24
; NUMBER OF SEQ ID NOS: 33136
```

```
; SEQ ID NO 17480
; LENGTH: 859
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; NAME/KEY: peptide
; LOCATION: 1..859
; OTHER INFORMATION: Ceres Seq. ID no. 3023746
US-09-935-625-17480
```

```
alignment_scores:
Quality: 210.50 Length: 380
Ratio: 1.126 Gaps: 23
Percent Similarity: 49.211 Percent Identity: 31.316
```

alignment\_block:

US-09-462-480-1/rev x US-09-935-625-17480 ..

Align seg 1/1 to: US-09-935-625-17480 from: 1 to: 859

```
1253 TCTCAGATCGCTCAACGCTATAACACGAGAAAGCGGACGCGG 1204
||||| ||| : : : : : ||| ||| ||| ||| ||| |||
266 SerArgSerProLleArgArg.....HisArgArgProThrHisGluL 280
||||| ||| : : : : : ||| ||| ||| ||| ||| |||
1203 AAGTCAACTCGCGCATCCGCTGTTTCGCTATCTACGGAACTCGG 1154
||||| ||| : : : : : ||| ||| ||| ||| ||| |||
280 yArgGlnSerProAlaPro.....SerArgArgArg 292
||||| ||| : : : : : ||| ||| ||| ||| ||| |||
1153 GTTCCCTATCGGAACATCCAGTGACGTTCCTTCGTCGCAACCATG 1104
||||| ||| : : : : : ||| ||| ||| ||| ||| |||
292 rgSerProSerProAlaArgArg.....ArgSerProSerPro 306
||||| ||| : : : : : ||| ||| ||| ||| ||| |||
1103 CCTGACCGGCTTCGCTGATCGTCGCGCCAGGTTCTCAGCGCTGTC 1054
||||| ||| : : : : : ||| ||| ||| ||| ||| |||
307 ProAlaArgArgArgSerProSerProAlaArgHisArgSe 323
||||| ||| : : : : : ||| ||| ||| ||| ||| |||
1053 AGTCGGTAGCGTGGCGTCCCATTTTGTGTGGACACCCCTGGTACGCTC 1004
||||| ||| : : : : : ||| ||| ||| ||| ||| |||
```

```
323 rProThrProProAlaArgGln.....ArgArgSerProSerProProA 338
1003 CGAACCGCTACCGG...CCGAGCGCGCTGGAGCTTGGTCAGGACTGCT 957
||||| ||| : : : : : ||| ||| ||| ||| ||| |||
338 laArgArgHisArgSerProProAlaArgArgArgArgSerProSer 354
||||| ||| : : : : : ||| ||| ||| ||| ||| |||
956 TCCCTCGTCAAGGAGGAATGAATGACGTGACATTTCCCTCGATTGCG 907
||||| ||| : : : : : ||| ||| ||| ||| ||| |||
355 ProProAlaArgArgArgArgSerProSerProProAlaArgArgArg 371
||||| ||| : : : : : ||| ||| ||| ||| ||| |||
906 CTTCGCGCGCGCTCGATACCGCGGAAATTCACCTGCTGCTCTCATGTT 857
||||| ||| : : : : : ||| ||| ||| ||| ||| |||
371 gSerProSerProLeuTyArgArgAsnArgSerPro..SerProLeuTy 387
||||| ||| : : : : : ||| ||| ||| ||| ||| |||
856 TTTCGCTCCGTTTCTTTTCGATTAGCGGTCAGAAAGCCCATTTTCG 807
||||| ||| : : : : : ||| ||| ||| ||| ||| |||
387 rArgArgAsn.....ArgSerArgSerProLeuAlaLysA 399
||||| ||| : : : : : ||| ||| ||| ||| ||| |||
806 CAGCGCTGCTGCTGCTCT...CGTCGCGCCTCGAGTATTGGACGCGG 760
||||| ||| : : : : : ||| ||| ||| ||| ||| |||
399 rgGlyArgSerAspSerProGlyArgSerProSerProValAlaArgLeu 415
||||| ||| : : : : : ||| ||| ||| ||| ||| |||
759 CTTGACCAATATTCGTCGAGA.....TCTCGTCGAGT 728
||||| ||| : : : : : ||| ||| ||| ||| ||| |||
416 ArgAspProThrGlyAlaArgLeuProSerProSerIleGluGlnArgLe 432
||||| ||| : : : : : ||| ||| ||| ||| ||| |||
727 TCTGCTTCTGCTTATTGGTCTTCTTGGAAAGCGCACCCG...CGGC 681
||||| ||| : : : : : ||| ||| ||| ||| ||| |||
432 uProSerProProValAlaGlnArgLeuProSerProProArgArgA 449
||||| ||| : : : : : ||| ||| ||| ||| ||| |||
680 CTGGCGCGCTCCCGCGCGCGCCACTGGCCCTGCAACAACTG 631
||||| ||| : : : : : ||| ||| ||| ||| ||| |||
449 laGlyLeuProSerProPro.....ProAlaGlnArgLeu 460
||||| ||| : : : : : ||| ||| ||| ||| ||| |||
630 CCGTCGACTCCACTGTCGATCTGGGTTTCAGGTCGCCGCGAGA..... 586
||||| ||| : : : : : ||| ||| ||| ||| ||| |||
461 ProSerProProProArgArgAlaGlyLeuProSerProMetArgIleG 477
||||| ||| : : : : : ||| ||| ||| ||| ||| |||
585 .....TCGCTCGAAATACCTG..... 568
||||| ||| : : : : : ||| ||| ||| ||| ||| |||
477 yGlySerHisAlaAlaAsnHisLeuGluSerProSerProSerLeuS 494
||||| ||| : : : : : ||| ||| ||| ||| ||| |||
567 ..CCTCTGCGCGAGGTAGCGCATCGTCTTCATCTGCCATGCTGG 520
||||| ||| : : : : : ||| ||| ||| ||| ||| |||
494 erProProGlyArg.....LysLysValLeuProSerProProValArg 508
||||| ||| : : : : : ||| ||| ||| ||| ||| |||
519 ACTACTTCTCTCTTACCTTCCTCGCCAAATGTTGCAAGTCTTCGG 470
||||| ||| : : : : : ||| ||| ||| ||| ||| |||
509 ArgArgArgSerLeuThrProAspGluGluArgValSerLeuSerGln.. 524
||||| ||| : : : : : ||| ||| ||| ||| ||| |||
469 CCGGTTGGCGGGAAGTCTGTTGCTATTAC.....GGGAGC. 433
||||| ||| : : : : : ||| ||| ||| ||| ||| |||
525 ....GlyGlyArgHisThrSerProSerHisIleLysGlnAspGlySerM 540
||||| ||| : : : : : ||| ||| ||| ||| ||| |||
432 ..TCACAGTCGCTCTTCGTCCTCCAGTCGCTCGTCTCTTCTTACG 385
||||| ||| : : : : : ||| ||| ||| ||| ||| |||
540 etSerProVal.ArgGlyArgGlyLysSerProSerSerArg.HisG 556
||||| ||| : : : : : ||| ||| ||| ||| ||| |||
384 CTCCTCGCGAGCGGTGCCGCGGACGACACCGG.....GGCTGG 344
||||| ||| : : : : : ||| ||| ||| ||| ||| |||
556 lnLysAlaArgSerProValArgArgArgSerProThrProValAsnArg 572
||||| ||| : : : : : ||| ||| ||| ||| ||| |||
343 TGGAGCGCGCGGATTCGAAACCTCGCTCCGCGGACCGACCGGA 294
||||| ||| : : : : : ||| ||| ||| ||| ||| |||
573 ArgSerArgArgSerSerSerAlaSerArgSer...ProAspArgArgA 588
||||| ||| : : : : : ||| ||| ||| ||| ||| |||
293 GCGGCGCCACCGG.....TCACCGACGATCCGCAACAGCGCG 256
||||| ||| : : : : : ||| ||| ||| ||| ||| |||
588 gArgArgSerProSerSerSerArgSerProSerArgSerArgSerProp 605
||||| ||| : : : : : ||| ||| ||| ||| ||| |||
255 CGGCA.....TCACCGAGGGGGCAA 236
||||| ||| : : : : : ||| ||| ||| ||| ||| |||
605 roValLeuHisArgSerProSerProArgGlyArg 616
```

seq\_name: /cgn2\_6/ptodata/2/paa/US09\_NEW\_COMB.pep:US-09-935-625-17479

```

seq_documentation_block:
; Sequence 17479, Application US/09935625
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: POLYNUCLEOTIDES, POLYPEPTIDES, CELLS, AND METHODS THEREOF CAPABLE
; OF INDUCING A SPECIFIC IMMUNE RESPONSE IN A HOST
; TITLE OF INVENTION: MODULATING VARIOUS RESPONSES
; FILE REFERENCE: 2750-1481P
; CURRENT APPLICATION NUMBER: US/09/935,625
; CURRENT FILING DATE: 2001-08-24
; NUMBER OF SEQ ID NOS: 33136
; SEQ ID NO 17479
; LENGTH: 879
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: peptide
; LOCATION: 1..879
; OTHER INFORMATION: Ceres Seq. ID no. 3023745
US-09-935-625-17479

```

alignment_scores:		
Quality:	210.50	Length: 380
Ratio:	1.126	Gaps: 23
Percent Similarity:	49.211	Percent Identity: 31.316

alignment\_block:

US-09-462-480-1/rev x US-09-935-625-17479

Align seq 1/1 to: US-09-935-625-17479 from: 1 to: 879

1253 TCTCAGATGCGCTCAAGAGTATAAACACAGAGAAGGGCGAGACCGCG 1204  
 286 SerArgSerProIleArgArg.....HisArgArgProThrHisGluG1 300  
 1203 AAGGTCGAACTCGCCGATCCCGTGTTCGCTATTCTACGCGAACATCGCG 1154  
 300 YArgGlnSerProAlaPro.....SerArgArgArgA 312  
 1153 GTTGCCCTATGCGAACATCCCAAGTCAGTGTGCTTCGTCGAAGCCATTG 1104  
 312 rgSerProSerProProAlaArgArg.....ArgSerProSerPro 326  
 1103 CTGACCGGCTTCGCTGATGTCGCGCGCAGGTTCTGACGCGGTGTTTC 1054  
 327 ProAlaArgArgArgSerProSerProProAlaArgArgHisArgSe 343  
 1053 AGCTCGGTAGCGGTGGCGTCCCATTTTCTGGACACCCCTGGTACGCCCTC 1004  
 343 rProThrProProAlaArgGln.....ArgArgSerProSerProProA 358  
 1003 CGAACCGGTACCGC...CCGAGCGCGCTCGAGCTGGTCAGGGAGCTGT 957  
 358 laArgArgHisArgSerProProProAlaArgArgArgSerProSer 374  
 956 TCCCGTCTCAAGGAGGAATGAATGGCGTACATTCCCTCGGATTCGCG 907  
 375 ProProAlaArgArgArgSerProSerProProAlaArgArgArgArg 391  
 906 CTTCGCGCGGCTCGATACCGCGAAATCCACTGCTGCTCTCTCATGTT 857  
 391 gSerProSerProLeuTyArgArgAsnArgSerPro..SerProLeuTy 407  
 856 TTTGCTCCGTTCTTTTCGTATTACGCGGTGAGAACCCCATTTTCGAGGA 807  
 407 rArgArgAsn.....ArgSerArgSerProLeuAlaLysA 419  
 806 CAGCGCGCTGCTGCTCTCT...CGTCGGCCCTCGAGTATTGGACGCCGG 760  
 419 rgGlyArgSerAspSerProGlyArgSerProSerProValaIaArgLeu 435

759 CCTGACGAATATTTCGTGGAGA.....TCTCGTCGAGT 728  
||| ::::||||  
436 ArgAspProThrGlyAlaArgLeuProSerProSerIleGluGlnArgLeu 452  
727 TCCTGCTTCTGCTATTGGCTGCTTCTTTGGNAGGCACCACG...CGCG 681  
:||||: ||| ||| ||||| |||  
452 uProSerProProValAlaGlnArgLeuProSerProProArgAra 469  
680 CTNGCGGGCGCTCCCGCGCGCGGCCACTGGCGCCTGCAACGAACCTGG 631  
||| ||||| ||||| .....ProAlaGlnArgLeu 480  
469 laGlyLeuProSerProPro..... ||||| :||| |||  
630 CGCTGCAGTCCACCTGGTCGATCTGGTGTTCAGGTCCGCCGAGA.... 586  
||||| ||||| ||||| :||| |||  
481 ProSerProProArgArgAlaGlyLeuProSerProMetArgIleGl 497  
585 .....TCGGCTCGAAATTACCTG..... 568  
||| ||||:||||  
497 yGlySerHisAlaAlaAsnHisLeuGluSerProSerProSerSerLeuS 514  
567 ..CCTCTCCCGGAGGTACGGCATCGGCTTTCATCTCTGCCATGCTGG 520  
|||||:|||| :||| ||| |||  
514 erProGlyArg.....LysLysValLeuProSerProProValArg 528  
519 ACTACTTTCTCTTTWACCTTCCTCGCAAATAATTGGCAAGTCTTCGGG 470  
||||| :|||:||||| |||||  
529 ArgArgArgSerLeuThrProAspGluGluArgValSerLeuSerGln.. 544  
469 CCCGGTGCCGGGAAGTCTGTTGTCATTAC.....GGGAGC. 433  
|||||:||||: |||  
545 ...GlyGlyArgHisThrSerProSerHisIleLysGlnAspGlySerM 560  
432 ..TCACCACTGCTCTTCGTCCTCCAGTCGTCCTGCTCTCTTCTCACG 385  
||| ||||| :||| :||| ||||| |||||  
560 etSerProVal.ArgGlyArgGlyLysSerProSerSerArg.Hisg 576  
384 CTCCTCGCGAGCGGTGCGCGCGCACGACCCG.....GGCTGG 344  
|||||: ||| ||| ||||| :|||  
576 InLysAlaArgSerProValArgArgSerProThrProValAsnArg 592  
343 TGGAGCCGCGGATTCCGAACCTGGCCCATCGTCCGGNACCCACCGGA 294  
||||| ||||| :||| :||| |||||  
593 ArgSerArgArgSerSerAlaSerArgSer...ProAspArgArgAr 608  
293 GCGGCGCCACCCG.....TCACGACGATCCGGCAACAGCCGC 256  
:||||| ||| |||||:||||  
608 gArgArgSerProSerSerSerArgSerProSerArgSerArgSerProp 625  
255 CGGCA.....TCACGAGGGGGCAA 236  
|||: ||||| ||||| |||||  
625 roValLeuHisArgSerProSerProArgGlyArg 636

seq\_name: /cqn2\_6/ptodata/2/paa/US09\_NEW\_COMB.pep:US-09-935-625-17075

seq documentation block:

```
seq_documentation_block:
  :; Sequence 17075, Application US/09935625
```

; GENERAL INFORMATION:

; APPLICANT: N. ALEXANDROV et al.

; TITLE OF INVENTION: POLYNUCLEOTIDES, POLYPEPTIDES

; TITLE OF INVENTION: MODUL

; FILE REFERENCE: 2750-1481P

; CURRENT APPLICATION NUMBER: US/0

; CURRENT FILING DATE: 2001-08-24

; NUMBER OF SEQ ID NOS: 331366

; SEQ ID NO 170

```
; LENGTH: 891
. TYPE: DPT
```

; TYPE: prt  
: ORGANTSM: Arabidopsis thaliana

```
;; ORGANISM: Arabidop
. FEATURE.
```

```

; FEATURE:
: NAME/KEY: peptide

```

```

; NAME/KEY: peplive
; LOCATION: 1-891

```

LOCATION: 1.691  
OTHER INFORMATION: Ceres Seq. ID no. 2708206

US-09-935-625-17075

## alignment\_scores:

Quality: 210.50 Length: 380  
Ratio: 1.126 Gaps: 23  
Percent Similarity: 49.211 Percent Identity: 31.316

## alignment\_block:

US-09-462-480-1/rev x US-09-935-625-17075

Align seg 1/1 to: US-09-935-625-17075 from: 1 to: 891

```
1253 TCTCAGAGTGGCTCAACAGCTATAACACAGAGAAAGCGGAGACCGAGCG 1204
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
298 SerArgSerProIleArgArg.....HisArgArgProThrHisGluG1 312
1203 AAGTGTGAACTCGCGGATCCCGTCCGTTTCGCTATTCTACGGAACCTGGC 1154
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
312 yArgArgGlnSerProAlaPro.....SerArgArgArgA 324
1153 GTTGCCCTATCGAATCCAGTGACGTTGCCTTCGGTCGAAGCCATTG 1104
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
324 IgSerProSerProAlaArgArg.....ArgSerProSerPro 338
1103 CTGACCGGCTTCGCTGATCGCGCGGAGTTCTGCAGCGCGTTGTT 1054
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
339 ProAlaArgArgArgSerProSerProProAlaArgArgHisArgSe 355
1053 AGTCGGTAGCGGTGGCGTCCCTTTTGTGGACACCTCGTACGCGCTC 1004
|||:|||||:|||||:|||||:|||||:|||||:|||||:
355 rProThrProProAlaArgGln.....ArgArgSerProSerProA 370
1003 CGAACCGCTACCGC...CCAGCGCGTCCGAGCTTGCTCAGGAGCTGCT 957
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
370 laArgArgHisArgSerProProAlaArgArgArgArgSerProSer 386
956 TCCCTCTCAAGGAGGAATGAATGACGTGCACATTTCCCTGGATTGCG 907
|||:|||||:|||||:|||||:|||||:|||||:|||||:
387 ProProAlaArgArgArgSerProSerProProAlaArgArgArgA 403
906 CTGCGCGGGCTCGATACCGCGGAATTCACATGCTGCTGTGATGTT 857
|||:|||||:|||||:|||||:|||||:|||||:|||||:
403 gSerProSerProLeuTyArgArgAsnArgSerPro..SerProLeuTy 419
856 TTGCTCGGTTTCTTTGCTATTAGCGGTGACAGGCCATTTGCGAGGA 807
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
419 rArgArgAsn.....ArgSerArgSerProLeuAlaLysA 431
806 CAGCGCTGCTGCTGCTCT...CGTCGGCCTCGAGTATTGGACGCGG 760
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
431 rgGlyArgSerAspSerProGlyArgSerProSerProValAlaArgLeu 447
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
759 CTGACGAATATTCGTCGAGA.....TCTCGTCGAGT 728
|||:|||||:|||||:|||||:|||||:|||||:|||||:
448 ArgAspProThrGlyAlaArgLeuProSerProSerIleGluGlnArgLe 464
727 TCCTGCTTCTGCTATTGCTGCTCTTGTGAAGCGCACCGG...CGGC 681
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
464 uProSerProProValAlaGlnArgLeuProSerProProArgArgA 481
680 CTGGCGCGCGTCCCGCGCGCGCCACTGGCGCTGCAACGAACTCGT 631
|||:|||||:|||||:|||||:|||||:|||||:|||||:
481 laGlyLeuProSerProPro.....ProAlaGlnArgLeu 492
630 CCGTCGACTCCACTGTGCTGATGTTGTTTCAGTCCGCCGAGAGA..... 586
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
493 ProSerProProProArgAlaGlnArgLeuProSerProMetArgIleG1 509
585 .....TCCGCTCGAATATTCCTG..... 568
|||:|||||:|||||:|||||:|||||:|||||:|||||:
509 yglySerHisAlaAlaAsnHisLeuGluSerProSerProSerSerLeus 526
```

```
567 ..CCTCTGCCCCGAGGAGTAGCGCATCGGTCTTCATCTCTGCCATGCTGG 520
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
526 erProProGlyArg.....LysLysValLeuProSerProProValArg 540
519 ACTACTTTCTCTTACCTTCCTCCGCCAAATGTTGGCAAGTCTTCGGG 470
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
541 ArgArgArgSerLeuThrProAspGluGluArgValSerLeuSerGln.. 556
469 CCCGGTGGCGCGGAAGTCTGTTGTCATTAC.....GGGAGC. 433
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
557 ...GlyGlyArgHisThrSerProSerHisIleLysGlnAspGlySerM 572
432 ..TCACCAAGTCTCTCTTCCTCCAGTCCTCGTCTGCTCTTCTTCACG 385
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
572 etSerProVal.ArgGlyArgGlyLysSerSerProSerSerArg.HisG 588
384 CTCCTGGCGAGCGGTGCCGCGGACGACGACCG.....GGCTGG 344
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
588 lnLysAlaArgSerProValArgArgArgSerProThrProValAsnArg 604
343 TGGAGCGCGCGGATTGGCAACCTCGCTCCGCGGACCCACCGGA 294
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
605 ArgSerArgArgSerSerSerAlaSerArgSer...ProAspArgArgA 620
293 GCGCGCGCACCG.....TCACCAAGTCTCGGCAACGCGGACGCGC 256
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
620 gArgArgSerProSerSerSerArgSerProSerArgSerArgSerProp 637
255 CGGCA.....TCACGAGGGGGCAA 236
:|||||:|||||:|||||:|||||:|||||:|||||:
637 roValLeuHisArgSerProSerProArgGlyArg 648
```

seq\_name: /cgn2\_6/ptodata/2/paa/US09\_NEW\_COMB.pep:US-09-935-625-17478

## seq\_documentation\_block:

; Sequence 17478, Application US/09935625

; GENERAL INFORMATION:

; APPLICANT: N. ALEXANDROV et al.

; TITLE OF INVENTION: POLYNUCLEOTIDES, POLYPEPTIDES, CELLS, AND METHODS THEREOF CAP.

; TITLE OF INVENTION: MODULATING VARIOUS RESPONSES

; FILE REFERENCE: 2750-1481P

; CURRENT APPLICATION NUMBER: US/09/935,625

; CURRENT FILING DATE: 2001-08-24

; NUMBER OF SEQ ID NOS: 33136

; SEQ ID NO 17478

; LENGTH: 903

; TYPE: PRT

; ORGANISM: Arabidopsis thaliana

; FEATURE:

; NAME/KEY: peptide

; LOCATION: 1..903

; OTHER INFORMATION: Ceres Seq. ID no. 3023744

US-09-935-625-17478

## alignment\_scores:

Quality: 210.50 Length: 380  
Ratio: 1.126 Gaps: 23  
Percent Similarity: 49.211 Percent Identity: 31.316

## alignment\_block:

US-09-462-480-1/rev x US-09-935-625-17478

Align seg 1/1 to: US-09-935-625-17478 from: 1 to: 903

```
1253 TCTCAGAGTGGCTCAACAGCTATAACACAGAGAAAGCGGAGACCGAGCG 1204
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
310 SerArgSerProIleArgArg.....HisArgArgProThrHisGluG1 324
1203 AAGTGTGAACTCGCGGATCCCGTTCGCTATTCTACGGAACCTCGGC 1154
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
324 yArgArgGlnSerProAlaPro.....SerArgArgArgA 336
```

```
1153 GTTGCCCTATGCGAACAATCCCAAGTACGCTTGGCTTCGGTCCGAAGCCATTG 1104
|| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
336 rgSerProSerProProAlaArgArg.....ArgSerProSerPro 350
1103 CTGACCGGCTTCGTGTATGCTCCCGCCAGGTTCCTGACGCGCTGTTTC 1054
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
351 ProAlaArgArgArgSerProProAlaArgArgHisArgSe 367
1053 AGTCGGTAGCGGTGCGTCCCATTTTGTGTCGACACCTGTCGCGCTC 1004
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
367 rProThrProProAlaArgGln.....ArgArgSerProSerProProA 382
1003 GCAACCGTACGCG...CCAGCGCGCTGCGAGCTTGTGTCAGGACTGCT 957
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
382 laArgArgHisArgSerProProAlaArgArgArgSerProSer 398
956 TCCCTCTCCTCAAGAGGAAGTAATGAGCGTCACATTTCCCTGGATTGCG 907
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
399 ProProAlaArgArgArgSerProSerProProAlaArgArgArg 415
906 CTTCGCGCGGCTCGATACCCCGGAAATTCACCTGCTGCTGTCATGTT 857
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
415 gSerProSerProLeuTyArgArgAsnArgSerPro...SerProLeuTy 431
856 TTTCGCTCCGTTCTTTCTGATTAGCGGGTCAGAACCCATTTCGAGGA 807
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
431 rArgArgAsn.....ArgSerArgSerProLeuAlaLysA 443
806 CAGCGCTCTGCTGCTCCT...CGTCGCGCTCGAGTATTGGAGCCGG 760
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
443 rgGlyArgSerAspSerProGlyArgSerProSerProValAlaArgLeu 459
759 CTGACGAATATTCTCGCAGA.....TCTCGTCGAGT 728
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
460 ArgAspProThrGlyAlaArgLeuProSerProSerIleGluGlnArgLe 476
727 TCTGCTCTCTGTTATTGCTGCTTCTTTGAAGCGCACCCG...CGCG 681
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
476 uProSerProProValAlaGlnArgLeuLeuProSerProProArgArg 493
680 CTGGCGCGCGTCCCGCGCGCGCCGCTGCGCTGCGCAACGACCTG 631
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
493 laGlyLeuProSerProPro.....ProAlaGlnArgLeu 504
630 CCGTCGACTCCACTGCTCGATCTGGTTTCAGGTCGCGCGAGA.... 586
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
505 ProSerProProProArgArgAlaGlyLeuProSerProMetArgIleG 521
585 .....TCCGCTCGAATTACCTG..... 568
521 yGlySerHisAlaAlaAsnHisLeuGluSerProSerProSerLeuS 538
567 ..CTTCCTCCCGAGGGTAGCGGATCGCTTTCATCTCTCCCATGCTGG 520
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
538 erProProGlyArg.....LysLysValLeuProSerProProValArg 552
519 ACTACTTCTCTTCTTACCTTCTCGCCCAAAATGTTGGCAAGTCTTCGG 470
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
553 ArgArgArgSerLeuThrProAspGluGluArgValSerLeuSerGln.. 568
469 CCCGGTGGCGGAGTCTGTTGTCATTAC.....GGGAGC. 433
569 ....GlyGlyArgHisThrSerProSerHisIleLysGlnAspGlySerM 584
432 ..TCACCACTGCTCTCTTCCTCCAGTCGCTGCTGCTCTTCTTCACG 385
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
584 etSerProVal..ArgGlyArgGlyLysSerSerProSerArg.HisG 600
384 CTCTGCGCGAGGTTGCGCGCGGACAGCACCGG.....GGTGG 344
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
600 lnLysAlaArgSerProValArgArgArgSerProThrProValAsnArg 616
343 TGGAGCCCGCGATTGCGAACCCCTGGCCCATCGCTCCCGGACCCACCGGA 294
```

```
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
617 ArgSerArgArgSerSerAlaSerArgSer...ProAspArgArgAr 632
293 GCGGCGCCACCCG.....TCACCGACGATCCGGCAACAGCCGC 256
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
632 gArgSerProSerSerArgSerProSerArgSerArgSerProp 649
255 CGGCA.....TCACCGAGGGGCAA 236
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
649 roValLeuHisArgSerProSerProArgGlyArg 660
seq_name: /cgn2_5/ptodata/2/paa/US09_NEW_COMB.pep:US-09-116-492A-7
seq_documentation_block:
; Sequence 7, Application US/09116492A
; GENERAL INFORMATION:
; APPLICANT: GICOUEL, BRIGITTE
; APPLICANT: BERTHET, FRANCOIS-XAVIER
; APPLICANT: ANDERSEN, PETER
; APPLICANT: RASMUSSEN, PETER B
; TITLE OF INVENTION: POLYNUCLEOTIDE FUNCTIONALLY CODING FOR THE LHP PROTEIN FROM MY
; TITLE OF INVENTION: TUBERCULOSIS, ITS BIOLOGICALLY ACTIVE DERIVATIVE FRAGMENTS, A
; TITLE OF INVENTION: USING THE SAME
; FILE REFERENCE: 0660-0137-27X
; CURRENT APPLICATION NUMBER: US/09/116,492A
; PRIOR FILING DATE: 1998-07-16
; PRIOR FILING DATE: 60/252,631
; PRIOR FILING DATE: 1997-07-16
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 42
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; US-09-116-492A-7
alignment_scores:
Quality: 205.00 Length: 42
Ratio: 4.881 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-462-480-1 x US-09-116-492A-7 ..
Align seg 1/1 to: US-09-116-492A-7 from: 1 to: 42
699 CAAGNAGCAGCCAATAACGAGCAGGAAGTCTGACGAGATCTCGACGAA 748
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1 GlnGluAlaAlaAsnLysGlnLysGlnLeuLeuAspGluIleSerThrAs 17
749 TATTGCTCAGCGCGCTCCCAATCTCGAGGCGCGCAGGAGCAGCAGCAGC 798
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
17 nIleArgGlnAlaGlyValGlnTySerArgAlaAspGluGlnGlnG 34
799 AGGCGCTGCTCTCGCAATGGGCTTC 824
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
34 lnAlaLeuSerSerGlnMetGlyPhe 42
seq_name: /cgn2_5/ptodata/2/paa/US10_NEW_COMB.pep:US-10-140-045-7
seq_documentation_block:
; Sequence 7, Application US/10140045
; GENERAL INFORMATION:
; APPLICANT: GICOUEL, BRIGITTE
; APPLICANT: BERTHET, FRANCOIS-XAVIER
; APPLICANT: ANDERSEN, PETER
; APPLICANT: RASMUSSEN, PETER B
; TITLE OF INVENTION: POLYNUCLEOTIDE FUNCTIONALLY CODING FOR THE LHP PROTEIN FROM MY
; TITLE OF INVENTION: TUBERCULOSIS, ITS BIOLOGICALLY ACTIVE DERIVATIVE FRAGMENTS, A
; TITLE OF INVENTION: USING THE SAME
; FILE REFERENCE: 0660-0137-27X
; CURRENT APPLICATION NUMBER: US/10/140,045
```

```
; CURRENT FILING DATE: 2002-05-08
; PRIOR APPLICATION NUMBER: US/09/116.492A
; FILING DATE: 1998-07-16
; PRIOR APPLICATION NUMBER: 60/252,631
; PRIOR FILING DATE: 1997-07-16
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 7
; LENGTH: 42
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-140-045-7
```

```
alignment_scores:
  Quality: 205.00      Length: 42
  Ratio: 4.881         Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
```

```
alignment_block:
US-09-462-480-1 x US-10-140-045-7
Align seg 1/1 to: US-10-140-045-7 from: 1 to: 42
```

```
699 CAAGAGCAGCCCAATAGCAGAGCAGGAACTCGACGAGATCTCGAGGAA 748
|||||
1 GlnGluAlaAlaAsnLysGlnLysGlnGluLeuAspGluIleSerThrAs 17
749 TATTCGTACAGCGCGCTCCAAATACCTCGAGGGCGGACGAGCAGCAGC 798
|||||
17 nileArgGlnAlaGlyValGlnTyrSerArgAlaAspGluGlnGlnG 34
799 AGCGCTGTCTCGCAATGGGCTTC 824
|||||
34 lnAlaLeuSerSerGlnMetGlyPhe 42
```

seq\_name: /cgn2\_6/ptodata/2/paa/US60\_NEW\_COMB.pep:US-60-389-987-2231

```
seq_documentation_block:
; Sequence 2231, Application US/60389987
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; FILE REFERENCE: 660088.465p2
; CURRENT APPLICATION NUMBER: US/60/389,987
; NUMBER OF SEQ ID NOS: 3025
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2231
; LENGTH: 2263
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-389-987-2231
```

```
alignment_scores:
  Quality: 203.50      Length: 442
  Ratio: 0.988         Gaps: 27
Percent Similarity: 46.606 Percent Identity: 28.733
```

```
alignment_block:
US-09-462-480-1 x US-60-389-987-2231
Align seg 1/1 to: US-60-389-987-2231 from: 1 to: 2263
```

```
33 GGGCGGCGCGGGCGGCGCAACCCAGCGAGGGA...AGCGCGCGCAGA 79
```

```
|||||
448 GlySerHisArgGluIleSerSerProThrSerLysAsnArgSerH1 464
80 TGGCGCTCGTCGCGCACCAGTCGCGTGCAGAACCATCGCTGGCTGGTGA 129
|||||
464 sGlyArgAlaLysArgAspLysSer.HisSerHisThr..... 476
130 TCAGGCCCCAGCGCGCGCGCTGCTCGCGCGAGTCGCTACCTGG 179
|||||
477 .....ProSerArgMetGlyArgSerArgSerProAlaThrAlaL 491
180 CGCAGGTGGGTGTCACCGCAGCGCGCTGATGTCACGTATCGGAAA 229
|||||
491 sArgGlyArgSerArgSerArgThrProThrLysArg.GlyHisSerArg 507
230 AGCGGTTGCCCTCGTGGTGATCGCGCGCTGTTGCCGGATCGTCGGTG 279
|||||
508 SerArgSerPro..... 511
280 ACGGCTGGCGCGCTCGGTGGTCCGG.....GAGCGATGGG 317
|||||
512 .....GlnTrpArgArgSerArgSerAlaGlnArgTrpG 523
318 CCAGGTTCCCAATCCG.....GCGGCTCCACAGCC 349
|||||
523 lyArgSerArgSerProGlnArgGlyArgSerArgSerProGlnArg 539
350 CGGCTGTGTCGCGCGCGCTCGCGCAGGCGTGAAGAGAGCAGC 399
|||||
540 ProGlyTrpSer.....ArgSerArgAsnThrGln.ArgArgGlyA 553
400 GAGGAGACTGGGACGAAGAGAGCAGCTGCTGAGTCCCTTAATGACAAC 449
|||||
553 rgSerArgSerAlaArgArgGly.....ArgSerHisSer 564
450 AGACTTCCCGCACCGCGCGGAGAACTTGCACAACTTTTGGCGGAGA 499
|||||
565 Arg.SerProAlaThrArg..GlyArg..... 572
500 AGGTAAGAGAGAAAGTAGTCCAGCATGGCAGAGATCAAGACCGATCCG 549
|||||
573 .....SerArgSerArgThrProAla..... 579
550 CTACCTCGCGCAGGAGGAGGTAATTTTCGAGCGATCTCGGCGACCTG 599
|||||
580 .....ArgArgGlyArgSerArgSerArgThrProAlaArgAr 592
600 AAAACCCAGATCGACAGGTGGAGTCGACGCGAGGTTGCTGCGAGGCCA 649
|||||
592 gArgSerArgSerArgThrProThrArgArgSerArgSerArgThr. 608
650 GTGGCGGCGCGCGGCGGAGCGCGCCAGCGCGCTGGTGGCTGCTCC 699
|||||
609 .....ProAlaArgArgGlyArg.....Ser 615
700 AAGAAGCAGCCAATAAGCAGAAGCAGGAAGCTCGACGAGATCTCGACGAAT 749
|||||
616 ArgSerArgThrProAlaArgArgSerArgSerArgThrArgSerPro..... 630
750 ATTCTGTCAGCGCGCTCCAAATCTCGAGGGCGGACGAGGAGCAGCAGA 799
|||||
631 ....ValArgArgSerArgSerArgSerProAlaArgSerGlyA 646
800 GGCCTGTCTCTCGCAATGGGCTTCTGACCCCGCTAATACGAAAAGAACG 849
|||||
646 rgSer.....ArgSerArgThr 651
850 GAGCAAAACATGACAGAGCAGCAGTGAATTTTCGGGGTATCGAGGCGG 899
|||||
652 ProAlaArgArgGlyArgSerArgSerArgThrProAlaArgArgGlyAr 668
900 CGGCAAG...CGCAATCCAGGGAATGTCACGTCATTCATTCCTCCTT 946
|||||
```





OM of: US-09-462-480-1 to: PIR\_71.\* out\_format : pfs  
Date: Jul 22, 2002 1:25 AM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=framet\_n2p\_model -DEV=xlh  
-O=/cgn2.1/USPTO.spool/US09462480/runat\_18072002\_164418\_19445/app\_query.fasta\_1.2850  
-DB=PIR\_71 -QFMT=fastan -SUFFIX=rpr -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -XGAPOP=4.500  
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000  
-DELEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZE=500  
-MINLEN=0 -MAXLEN=2000000000 -USER=US09462480.@CGN1\_1\_73  
-NCPU=6 -ICPU=3 -LONGLOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30  
-NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-462-480-1  
Query length: 1277  
Database: PIR\_71.\*  
Database sequences: 283138  
Database length: 96089334  
Search time (sec): 87.620000

score\_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
PIR2:G70802	+	713.00	852.17	6.1e-40	368	! Probable PPE protein - Mycobact
PIR2:H70802	+	486.00	591.14	7.7e-25	100	! hypothetical protein Rv3874 - M
PIR2:A70803	+	479.00	583.18	2.3e-24	95	! early secretory antigen target e
PIR1:Q08E3	-	212.50	248.07	1.5e-06	660	! BHLF1 protein - human herpesvir
PIR2:G84693	-	210.50	243.19	2.1e-06	391	! probable proline-rich protein
PIR2:T10033	-	198.00	237.17	1.3e-05	302	! hypothetical protein MLCB628.14
PIR2:T02345	+	195.50	219.46	2.2e-05	1791	! hypothetical protein KIAA0324
PIR2:T10032	+	187.00	232.13	6.8e-05	100	! hypothetical protein MLCB628.13
PIR2:T10031	+	186.00	232.36	7.9e-05	95	! early secretory antigen target e
PIR2:T13078	-	185.00	213.84	0.0001	772	! KIAA0992 protein - human (fragm
PIR2:S52796	-	182.50	216.22	0.0001	403	! prpL2 protein - human (fragm
PIR2:T02345	-	179.50	200.30	0.0003	1791	! hypothetical protein KIAA0324
PIR2:B40505	+	178.50	198.37	0.0003	1958	! hypothetical protein - suid he
PIR2:T08179	+	176.00	204.62	0.0004	640	! LRG5 protein - chlamydomonas re
PIR2:A45748	+	174.00	199.22	0.0006	920	! collagen alpha 1(VII) chain - m
PIR2:J00405	+	172.50	195.91	0.0007	1106	! hypothetical 119.5K protein (u
PIR1:B45344	+	172.50	192.19	0.0007	1733	! probable nuclear antigen - su
PIR2:S27923	+	171.00	195.60	0.0009	924	! gene Lr3 protein - human herpes
PIR2:S28774	+	171.00	194.72	0.0009	1027	! collagen alpha chain - tube wd
PIR1:EDBEIF	+	170.50	191.22	0.0010	1460	! immediate-early protein IE180
PIR2:F75311	+	167.00	195.06	0.0016	552	! ABC transporter, ATP-binding pr
PIR2:S08032	-	165.00	186.36	0.0023	1184	! atrophin-1 - human
PIR2:G01763	-	164.50	185.76	0.0025	1184	! atrophin-1 - human
PIR2:T31611	+	164.50	183.35	0.0025	1585	! hypothetical protein Y508A.g
PIR2:A54849	+	164.00	177.64	0.0028	2944	! collagen alpha 1(VII) chain pr
PIR2:A59266	-	163.50	175.54	0.0031	3530	! unconventional myosin-15 - hum
PIR2:T01696	-	162.50	191.82	0.0032	426	! pistil extensin-like protein ph
PIR2:T04859	-	162.50	186.22	0.0033	839	! extensin homolog F28A21.80 - A
PIR2:S41342	+	161.00	184.64	0.0042	817	! verprolin - yeast (Saccharomyce
PIR2:IA48103	+	160.00	178.15	0.0050	1549	! type VII collagen - Chinese ha
PIR1:A43291	+	159.50	178.55	0.0054	1373	! collagen alpha 2(I) chain pred
PIR2:G84348	+	159.00	186.40	0.0055	494	! hypothetical protein Vng1983h
PIR2:T43556	-	158.50	184.56	0.0060	574	! Wiskott-Aldrich syndrome protei
PIR2:T38819	-	158.50	184.56	0.0060	574	! Wiskott-Aldrich syndrome protei
PIR1:A45344	-	158.50	176.93	0.0063	1446	! immediate-early protein - suid
PIR1:CGHU2V	+	158.50	176.65	0.0063	1496	! collagen alpha 2(V) chain pred
PIR2:T45134	+	157.50	184.04	0.0069	529	! hypothetical protein [imported]
PIR1:EDBEXD	+	157.50	180.37	0.0071	825	! immediate-early protein RL2 - H
PIR2:A36068	+	157.00	180.92	0.0076	718	! major ampullate fibroin protein
PIR2:T05352	+	157.00	179.46	0.0077	857	! hypothetical protein F8B4.120 -

PIR2:G84693 + 157.00 179.13 0.0077 891 ! probable proline-rich protei  
PIR1:A45344 + 156.50 174.53 0.0086 1416 ! immediate-early protein - s  
PIR2:T46289 + 156.00 178.21 0.0090 862 ! hypothetical protein DKF2p4:  
PIR1:CGCH2S + 156.00 177.29 0.0091 964 ! collagen alpha 2(I) chain p  
PIR1:CGCH1S + 156.00 176.64 0.0091 1042 ! collagen alpha 1(I) chain  
seq\_name: pir2:G70802  
seq\_documentation\_block:  
Probable PPE protein - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
C:Accession: G70802  
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon  
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete gen  
A:Reference number: A70500; MUID:98295987  
A:Accession: G70802  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-368 <COL>  
A:Cross-references: GB:AL123456; GB:AL123456; NID:g3261558; PIDN:CAA17965.1; PID:e12  
A:Experimental source: strain H37RV  
C:Genetics:  
A:Gene: PPE

alignment\_scores:  
Quality: 713.00 Length: 143  
Ratio: 5.057 Gaps: 0  
Percent Similarity: 98.601 Percent Identity: 96.503  
alignment\_block:  
US-09-462-480-1 x G70802 ..  
Align seg 1/1 to: G70802 from: 1 to: 368

1 CTGACGAGGTGACGCTGTTGTTTTCAGCCAGGTGGCGGCACCGCGGGCGG 50  
|||||  
226 LeuGlnGlnValThrSerLeuPheSerGlnValGlyGlyThrGlyGlyG 242  
51 CAACCCACGCGCACGAGGAGCGCGCACATGGCTGCTGCGCACACGTC 100  
242 YAsnProAlaaspGluGluAlaGlnMetGlyLeuLeuGlyThrSerP 259  
101 CGCTGTCGAACATCCGCTGGCTGGTGGATCAGGCCCGCGCGCGCGCG 150  
259 roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla 275  
151 GGCCTGCTGCGCGGGAGTCCCTACCTGGCGGCGAGGTGGTGGTTCACCCG 200  
|||||  
276 GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr 292  
201 CACGCGCGTGTGCTCAGCTGATCGAAAGCGGTTGCCCTCCCTCGGTGA 250  
292 gThrProLeuMetSerGlnLeuLeuGluLysProValAlaProSerValM 309  
251 TCCCGCGCGGTGTTCCCGGATCGTGGTGGTGGCGGGTGGCGCGCTCCG 300  
309 etProAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325  
301 GGTCCGGAGGATGGCGCGGGTTCGGAATCCGCGCGCTCCACCGCCC 350  
|||  
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342  
351 GGGTCTGCTGCGCGCGCACCGCTCGCGCAGGAGGTGAAGAAGACGACG 400  
342 oGlyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluaspasp 359  
401 AGGACGACTGGGACGACGAGGACGACTGG 429

359 luAspAspTrpAspGluAspAspTrp 368  
|||||

seq\_name: pir2:H70802

seq\_documentation\_block:  
hypothetical protein RV3874 - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C:Accession: H70802  
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; MUID:98295987  
A:Accession: H70802  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-100 <COL>  
A:Cross-references: GB:AL022120; GB:AL123456; NID:g3261558; PIDN:CAAL7966.1; PID:g296022  
A:Experimental source: strain H37RV  
C:Genetics:  
A:Gene: RV3874

alignment\_scores:  
Quality: 486.00 Length: 100  
Ratio: 4.860 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 99.000  
alignment\_block:  
US-09-462-480-1 x H70802 ..  
Align seg 1/1 to: H70802 from: 1 to: 100

525 ATGGCAGAGATGACAGACCGATGCGCTACCTCGGCGAGGAGGAGGTA 574  
|||||  
1 MetAlaGluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlys 17  
575 TTTCAGCGGATCCCGCGACCTGAAACCCAGATCGACCATGGTGGAGT 624  
|||||  
17 nPheGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGluS 34  
625 CGAGCGGAGTTCGTTGAGCGCCAGTGGCGCGCGCGCGGAGGAGGCC 674  
|||||  
34 erThrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAla 50  
675 GCCCAGCGCGCGGTGGTGGCTTCCAGAGACGCCCAATAAGCAGAGCA 724  
|||||  
51 AlaGlnAlaAlaValValArgPheGlnGluAlaAlaAsnLysGlnLysG 67  
725 GGAACCTGACGAGATCGCAGCATATTCGTCAGCGCGGCTCCAACTACT 774  
|||||  
67 nGluLeuAspGluIleSerThrAsnIleArgGlnAlaGlyValGlnTrpS 84  
775 CGAGCGCGCGAGGAGCAGCAGCAGCGGCTGCTCCGCAATGGGCTTC 824  
|||||  
84 erArgAlaAspGluGlnGlnGlnGlnAlaLeuSerSerGlnMetGlyphe 100

seq\_name: pir2:A70803

seq\_documentation\_block:  
early secretory antigen target esat6 - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
C:Accession: A70803; S49174  
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno  
A:Reference number: A70500; MUID:98295987  
A:Accession: A70803  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-95 <COL>  
A:Cross-references: GB:AL022120; GB:AL123456; NID:g3261558; PIDN:CAAL7967.1; PID:e126  
A:Experimental source: strain H37RV  
R:Soerensen, A.L.; Nagai, S.; Houen, G.; Andersen, P.; Andersen, A.B.  
submitted to the EMBL Data Library, June 1994  
A:Reference number: S49174  
A:Accession: S49174  
A:Molecule type: DNA  
A:Residues: 1-13,'R',15-22,'S',24-95 <SOE>  
A:Cross-references: EMBL:X79562; NID:g531708  
C:Genetics:  
A:Gene: esat6

alignment\_scores:  
Quality: 479.00 Length: 95  
Ratio: 5.042 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-09-462-480-1 x A70803 ..  
Align seg 1/1 to: A70803 from: 1 to: 95

860 ATGACAGACGACGAGTGGAAATTCGCGGGTATCGAGCGCGGCGCAAGCGC 909  
|||||  
1 MetThrGluGlnGlnTrpAsnPheAlaGlyIleGluAlaAlaAlaSerAl 17  
910 AATCCAGGGAATGTCACGTCCATTTCCTCTTCGACGAGGGAAGC 959  
|||||  
17 alleGlnGlyAsnValThrSerIleHisSerLeuLeuAspGluGlyLysG 34  
960 AGTCCCTGACCAAGCTCGACGCGCTGGCGGTAGCGGTTTCGGAGCGG 1009  
|||||  
34 InSerLeuThrLysLeuAlaAlaAlaTrpGlyGlySerGlySerGluAla 50  
1010 TACACGGGTGTCACGCAAAATGGACGCCGCTACCGAGCTGAACAA 1059  
|||||  
51 TyrGlnGlyValGlnGlnLysTrpAspAlaThrAlaThrGluLeuAsnAs 67  
1060 CGCGTGCAGAACCTGGCGCGACGATCAGCGAGCGGTCAGGCAATGG 1109  
|||||  
67 nAlaLeuGlnAsnLeuAlaArgThrIleSerGluAlaGlyGlnAlaMetA 84  
1110 CTTGACCGAAGCAACGTCACCTGGGATGTTTCGCA 1144  
|||||  
84 laSerThrGluGlyAsnValThrGlyMetPheAla 95

seq\_name: pir1:Q0BE3

seq\_documentation\_block:  
BHLF1 protein - human herpesvirus 4 (strain B95-8)  
C:Species: human herpesvirus 4, Epstein-Barr virus  
C>Date: 25-Feb-1985 #sequence\_revision 25-Feb-1985 #text\_change 23-Aug-1997  
C:Accession: A03742  
R:Bankier, A.T.; Deininger, P.L.; Farrell, P.J.; Barrell, B.G.  
Mol. Biol. Med. 1, 21-45, 1983  
A:Title: Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8 Epstein-Barr vi  
A:Reference number: A93065; MUID:85035713  
A:Accession: A03742  
A:Molecule type: DNA  
A:Residues: 1-660 <BAN>  
R:Baer, R.; Bankier, A.T.; Biggin, M.D.; Deininger, P.L.; Farrell, P.J.; Gibson, T.J.  
Nature 310, 207-211, 1984  
A:Title: DNA sequence and expression of the B95-8 Epstein-Barr virus genome.  
A:Reference number: A03794; MUID:84270667  
C:Contents: annotation; protein coding region  
C:Comment: The sequence contains four perfect repeats (residues 149-273, 274-398, 399

C:Superfamily: human herpesvirus 4 BHLF1 protein

## alignment\_scores:

Quality: 212.50 Length: 560  
Ratio: 0.944 Gaps: 29  
Percent Similarity: 40.179 Percent Identity: 25.000

## alignment\_block:

US-09-462-480-1/rev x Q0BE3

Align seg 1/1 to: Q0BE3 from: 1 to: 660

```
1202 AGTGAACATCGCCGATCCCGTGT...CGCTATTCTACGGAAC 1159
||||| ||| ||||| ||||| ||||| ||||| ||||| |||||
136 ArgGlyProArgProHisProAlaPheGlnValGlnTrpSerAlaArgAs 152
1158 TCGCGGTGGCTATGCGACATCCAGTACGCTTGCCTTCGGTCGAAGC 1109
: ||||| ||||| ||||| ||||| ||||| ||||| |||||
152 nProGlyCysPro...ArgThrTrpArgArgSerGlyAlaGlnArgG 168
1108 CATTCGCTGACCGCTTCGCTGATCGTCGCGCCAGGCTTCGACGCGCT 1059
||| ||||| ||||| ||||| ||||| ||||| |||||
168 lyHisPro.....ProProGlyAlaGlyGlnArg 177
1058 TGTTCAGCTCGTAGCGGTGGCTGCCATTTTGTGGACACCCCTGTGAC 1009
||||| ||||| ||||| ||||| ||||| ||||| |||||
178 ProSerGlyProThrGlyGlyArgProAlaAlaProGlyAlaProGlyTh 194
1008 GCCT.....CCGAACCGCTACC 992
||| ||||| ||||| ||||| ||||| ||||| |||||
194 rProAlaAlaProGlyProGlyGlyAlaAlaValProSerGlyAlaT 211
991 GCCCGAGCGCTCGAGCTGTGTCAGGACTGCTCCCTCGTCFCAAGGA 942
||||| ||||| ||||| ||||| ||||| ||||| |||||
211 hrProHisProGluArgGly...SerGlyProAlaAspProProAlaAla 226
941 GGAATGAATGGAGTGCATATTCCTCGGATTCGCTGCGG..... 900
: ||||| ||||| ||||| ||||| ||||| |||||
227 Ala.....ArgLeuProProGluAr 233
899 ....CGCCCTCGATACCGGGAATTCCTGCTGCTGCTGCTGCTGCTTT 854
||||| ||||| ||||| ||||| ||||| ||||| |||||
233 gGlnGluProArgLeuProGlnAspLeuAlaAlaGlnArgCysProA 250
853 GTCCTGTTCTTTTCGATTAGCGGTCAGAACCCCATTT. GCGAGGACA 805
: ||||| ||||| ||||| ||||| ||||| ||||| |||||
250 laGlyProProProThrArgSerGlyAlaAlaAlaGlnArgThrHisArg 266
804 GCGCTGCTGCTGCTCCTCGTCGG..... 781
||| ||||| ||||| ||||| ||||| ||||| |||||
267 ArgProProGlyCysProArgSerAlaArgAsnProGlyCysProArgTh 283
780 .....CCCTCGAGTATG 768
||| ||||| ||||| ||||| ||||| ||||| |||||
283 rTrpArgArgSerGlyAlaGlnArgGlyHisProProGlyAlaG 300
767 GACCGCGCTGAGATATTCGCGAGATCTCGTCGAGTTCCTGCTTCT 718
||||| ||||| ||||| ||||| ||||| ||||| |||||
300 lyGlnArgPro.....SerGlyProThrGlyGlyArgProAlaAla 313
717 GCTTATTGGCTGCTTCTTGAAGCGCACCCGCGGCTGGCGCGCTC 668
||||| ||||| ||||| ||||| ||||| ||||| |||||
314 Pro.....GlyAlaProGlyThrProAlaAlaProGI 324
667 CCGG.....CCGCGCC 657
||| ||||| ||||| ||||| ||||| ||||| |||||
324 yProGlyGlyGlyAlaAlaValProSerGlyAlaThrProHisProGluA 341
656 GCGCCACTGGCCCT.....GCAAGCACTGGCG..... 628
||||| ||||| ||||| ||||| ||||| ||||| |||||
341 rgGlySerGlyProAlaAspProProAlaAlaAlaArgLeuProProGlu 357
```

```
628 ..... 628
358 ArgGlnGluProArgLeuProGlnAspLeuAlaAlaAlaGlnArgCysPr 374
627 ....TCGACTCCACCTGGTGCATCTGGTGTTCAGTTCGCGCGG..... 589
: ||||| ||||| ||||| ||||| ||||| ||||| |||||
374 oAlaGlyProProProThrArgSerGlyAlaAlaAlaGlnArgThrHisA 391
588 .....AGATCCGCTCGAAT..... 574
||||| ||||| ||||| ||||| ||||| ||||| |||||
391 rgArgProProGlyCysProArgSerAlaArgAsnProGlyCysProArg 407
573 .....TACCTGCTCCTGCGCG 557
: ||||| ||||| ||||| ||||| ||||| |||||
408 ThrTrpArgArgArgSerGlyAlaGlnArgGlyHisProProGlyAl 424
556 AGGTAGCGGATCGGTCTTCATCTCGCCATGCTGGACTACTTCTCTC 507
||| ||||| ||||| ||||| ||||| ||||| |||||
424 aGlyGlnArg.....ProSerGlyProThrGlyGlyA 435
506 TTTACCTTCCTCGCCAAATGTTGGCAAGTCTCCGCGCGCGGT...GGC 460
||| ||||| ||||| ||||| ||||| ||||| |||||
435 rgProAlaAlaProGlyAlaProGlyThrProAlaAlaProGlyProGly 451
459 CGGAAGTCTGTTGTCATTACGGAGCT.....CACCAGTCGCTCTT 416
||| ||||| ||||| ||||| ||||| ||||| |||||
452 .GlyGlyAlaAlaValProSerGlyAlaThrProHisProGluArgGlys 468
415 GGTCCAGTCTGCTCTGCTCTTCTTCACGC..... 384
||||| ||||| ||||| ||||| ||||| ||||| |||||
468 erGlyProAlaAlaAspProAlaAlaAlaArgLeuProProGluArgGln 484
383 .....TCCTGCGCGAGCGG 370
||| ||||| ||||| ||||| ||||| ||||| |||||
485 GluProArgLeuProGlnAspLeuAlaAlaAlaGlnArgCysProAlaGl 501
369 TCGCGCGCGACACCGCGGTGGTG.....GAGC 338
||| ||||| ||||| ||||| ||||| ||||| |||||
501 yProProProThrArgSerGlyAlaAlaAlaGlnArgThrHisArgArgp 518
337 GCGCGGATTCGGAACCTCGCCCATCTCCGCGACCCACCG..... 296
||||| ||||| ||||| ||||| ||||| ||||| |||||
518 roProGlyCysProArgSerAlaArgAsnProGly.CysProArgThrTr 534
295 .....GAGCGCGCCACCGCTCACCGACGATCCCGG 265
534 pArgArgSerGlyAlaGlnArgGlyHisProProProGlyAlaGlyG 551
264 AACAGCGCGCATCACCGAGGGGCAACCGGCTTTTCGATCAGCTCAG 215
||||| ||||| ||||| ||||| ||||| ||||| |||||
551 lnArgProSerGlyProThrGlyArgProAlaAlaProGlyAlaPro 567
214 ACATCAGCGGGTGC.....GGTCAACGACCCACCTGCGCCAGGT 174
: ||||| ||||| ||||| ||||| ||||| ||||| |||||
568 GlyThrProAlaAlaProGlyProGlyGlyAlaAlaValProSerGl 584
173 AGCGACTCCG.....CGCGCAGCAGCGCGG..... 149
||||| ||||| ||||| ||||| ||||| ||||| |||||
584 yAlaThrProHisProGluArgGlySerGlyProAlaAspProProAla 601
148 .....CGCCGCGCTGGGCGCTGATCCAC 125
: ||||| ||||| ||||| ||||| ||||| ||||| |||||
601 laAlaArgLeuProProGluArgGlnGluProArg..... 612
124 CAGCCAGCGATGTTTCAGACGCGACTGGTGGCAGCAGGCCCA..... 80
||| ||||| ||||| ||||| ||||| ||||| |||||
613 LeuProGlnAspLeuAlaAlaAlaGlnArgCysProAlaGlyProProPr 629
79 .....TCTGCGCGCTTCTCTCTCGG...CTGGGTT 52
oThrArgSerGlyAlaAlaAlaGlnArgThrHisArgArgProProGlyC 646
51 GCCCGCGCGGTGCGCGCCACCTGGC 26
```



```
alignment_scores:
  Quality: 198.00      Length: 126
  Ratio: 2.329        Gaps: 4
  Percent Similarity: 67.460  Percent Identity: 41.270

alignment_block:
US-09-462-480-1 x T10033  ..
Align seg 1/1 to: T10033 from: 1 to: 302

1  CTCGACGAGTGCCTGTTTCAGCCAGTGGCGGCACCGCGCGG 50
   :::::::::::::::::::::  :::::::::::::::::::::
190 MetGlnGlnValysSerLeuPheThrSerIleAspSerThrGlyValty 206
   :::::::::::::::::::::  GAGGAAGCCGCGCAGATGGCCTGC 88
   :::::::::::::::::::::  :::::::::::::::::::::
206 rThrSerAlaGlnArgGlyAspThrGluSerAlaHisArgIleGlyLeup 223
   :::::::::::::::::::::  :::::::::::::::::::::
89  TCGGCACACAGTCCGCTGTGCAACCATCCGCTGGCTGGATCAGGCC 138
   :::::::::::::::::::::  :::::::::::::::::::::
223 heGlyAlaSerThrLeuSerSerHisProLeuValGlyIleThrGlyThr 239
   :::::::::::::::::::::  :::::::::::::::::::::
139 AGCGCGGCGCGGCGCTGCTGCGCGGAGTCCGTACCTGCGCAGAGTGG 188
   :::::::::::::::::::::  :::::::::::::::::::::
240 ThrThrAspThrArgLeuLeuCysAlaGluSerLeuProSerAlaSerG1 256
   :::::::::::::::::::::  :::::::::::::::::::::
189 GTCGTTGACCCGCGCGCGCTGATGCTCT.....CAGCTGATCAAAAGC 232
   :::::::::::::::::::::  :::::::::::::::::::::
256 ySerLeuAlaTrpThrProLeuMetThrGlnPheGlnLeuIleAspLys 273
   :::::::::::::::::::::  :::::::::::::::::::::
233 CGGTGTCGCCCTCG.....GTGATCGCGCGCGCTGTTGCC 267
   :::::::::::::::::::::  :::::::::::::::::::::
273 erIleAlaProGluProArgGlnArgValMetLeuProTrpAlaAla 289
   :::::::::::::::::::::  :::::::::::::::::::::
268 GGATCGTGTGTTACGGTGGCGCGCTCCGCTGGGTCCGGAGCGATGGG 317
   :::::::::::::::::::::  :::::::::::::::::::::
290 GlySerPro.....G.....G1 293

318 CCAGGTTGCGCAATCCGCGCTCCACC 345
   :::::::::::::::::::::  :::::::::::::::::::::
293 yHisAsnAlaGlnAspGlyGlyThr 302

.seq_name: pir2:T02345

seq_documentation_block:
hypothetical protein KIAA0324 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 05-Nov-1999
C:Accession: T02345
R:Ricke, D.O.; Bruce, D.; Mundt, M.; Doggett, N.; Munk, C.; Saunders, E.; Robinson, D.;
re, J.; White, S.; Ueng, S.; Tatum, O.; Campbell, C.; Fawcett, J.; Deaven, L.
submitted to the EMBL Data Library, March 1998
A:Description: Sequencing of human chromosome 16p13.3.
A:Reference number: Z14664
A:Accession: T02345
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1791 <RIC>
A:Cross-references: EMBL:AC004493; NID:g2996648; PIDN:AAC08453.1; PID:g2996650
C:Genetics:
A:Map position: 16
A:Introns: 1610/2; 1706/2
A>Note: KIAA0324

alignment_scores:
  Quality: 195.50      Length: 474
  Ratio: 0.846        Gaps: 24
  Percent Similarity: 48.734  Percent Identity: 27.215

alignment_block:
US-09-462-480-1 x T02345  ..
Align seg 1/1 to: T02345 from: 1 to: 1791

48  CGGCAACCCAGCCGACGAGGAGCGCGCAGATGGGCT.....GC 88
   ||| ||||| ||||| :::::::::::::::::::::
465 ArgThrProSerArgArgSerArgSerGlySerProGlyLeuArgAs 481
   ||||| :::::::::::::::::::::
89  TCGGCACCACTCCGCTGTCAACCATCCGCTGGCTGGTGGATCA..... 132
   ||||| :::::::::::::::::::::
481 p.GlySerGlyThrProSerArgHisSerLeuSerGlySerProGly 497
   ||||| :::::::::::::::::::::
133 .....GGCCCCAGCGCGCGCGCGCTGCTGCTCGCGCGC 164
   ||||| ||| :::::::::::::::::::::
498 MetLysAspIleProArgThrProSerArgGly.....ArgSe 510
   :::::::::::::::::::::  TCGCTACCTGGCGCAGGTGGGT 190
   ||| :::::::::::::::::::::
510 rGluCysAspSerSerProGluProLysAlaLeuProGlnThrProArgp 527
   ||| CGTTGACCCGCGCGCTGATGCTCAGCTGATCGAAAAAGCGGTGCC 240
   :::::::::::::::::::::  :::::::::::::::::::::
527 roArgSerArgSerProSerProGluLeuAsnAsnLysCysLeuThr 543
   :::::::::::::::::::::  :::::::::::::::::::::
241 CCCTCGGTGATGCGCGCGCTGTTGCGCGATCGTGGGTGACGGGTGGCGC 290
   ||||| :::::::::::::::::::::
544 ProGln.....ArgGluArgSerGlySerGluSerValAspG1 557
   :::::::::::::::::::::  :::::::::::::::::::::
291 CGTCCCGTGGTCCGGGAGCGATGGGCGCAGGTTCGCAATCCGCGGCT 340
   ||||| :::::::::::::::::::::
557 nLysThrValAlaArgThrProLeuGlyGlnArgSerArgSerGlySer 574
   ||||| :::::::::::::::::::::
341 CCACCGCGCGGTCTGCTGCGCGCGCACCGCTCGCGCAGGAGCGTAA 390
   ||||| :::::::::::::::::::::
574 erGlnGluLeuAsp...ValLysProSerAlaSerProGlnGlnArgSer 589
   ||||| ||| ||||| :::::::::::::::::::::
391 GAAGACGACGAGGAGCGACTGGGACGAGGAGGACGACTGGTGGAGCTCC 440
   ||||| :::::::::::::::::::::
590 GluSer.....AspSerSerProAspSerLysAlaLysTh 601
   :::::::::::::::::::::  :::::::::::::::::::::
441 ANTGACACAGACTTCCCGCGCACCGCGCGGAGAGACTTGCACACATT. 489
   : ||| :::::::::::::::::::::
601 rArgThrProLeuArgGlnArgSerArgSerGlySerProGluValA 618
   :::::::::::::::::::::  TGGCGAGGAGGTAAAGAGAGAAAGTAGTCCACGA 525
   ||||| :::::::::::::::::::::
618 spSerLysSerArgLeuSerProArgArgSerArgSerGlySerPro 634
   ||| TGGCAGAGATGAAGACCGCTACCGCTACCGCGCGCAGGCGCAGTAAT 575
   :::::::::::::::::::::  ||| :::::::::::::::::::::
635 GluValLysAspLysProArgAlaProArgAlaGlnSerGlySerAs 651
   ||| TTCAGAGCGGATCTCGGCGCAGCTGAAACCCAGATCG...ACCAGTGA 622
   ||||| :::::::::::::::::::::
651 pSerSerProGluProLysAlaProAlaProArgAlaLeuProArgS 668
   ||||| :::::::::::::::::::::
623 GTCGACGCGCAGTTCGTTGACGGCCAGTGGCGCGCGCGCGGAGCGG 672
   ||||| :::::::::::::::::::::
668 erArg.....SerGlySerSerLysGlyArg 677
   :::::::::::::::::::::  :::::::::::::::::::::
673 CCGCCGCGCGCGGTGGTGGCTTCCAAAGAG.....CAGCCCAAT 713
   ||| ||||| :::::::::::::::::::::
678 GlyProSerProGluGlySerSerSerThrGluSerSerProGluHisPr 694
   ||| AAGCAAGACGAGGAACCTCGACGAGATCTCGACGAATATTCGTACGCGCG 763
   :::::::::::::::::::::  ||||| :::::::::::::::::::::
694 oProLysSerArgThrAlaArgArgGlySerArgSerProGluProL 711
   ||||| CGTCCAATACTGA...GGCGCGCAGGAGCAGCAGCGCGCTGCTCT 810
   :::::::::::::::::::::  ||| ||| ||||| |||||
711 ysThrLysSerArgThrProProArgArgArgSerSerArgSerPro 727
   :::::::::::::::::::::  CGCAATGGGCTTC 824
```



```

alignment_scores:
  Quality: 182.50      Length: 455
  Ratio: 1.031         Gaps: 28
  Percent Similarity: 38.901      Percent Identity: 27.473

alignment_block:
  US-09-462-480-1/rev x S52796 ..

```

Align seg 1/1 to: S52796 from: 1 to: 403

```
1268 GCGGCATGACAACTCTCAGAGTCGCTCAAAACGTATATAACACGAGAAA 1219
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
15 AlaGlyMetProLysLeuArgSerThrAlaAsnArgAspAsnAspSerG1 31
1218 GGGCGAG.....ACCGACGGAAGGTGCAACTGCGCCG 1187
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
31 yGlySerArgProProLeuLeuProProGlyGlyArgSerThrSerAlaL 48
1186 ATCCCGTGTTCGCTATTCTAGCGAACTCGCGTTCGCCCTATGGGACAA 1137
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
48 ysPro...PheSerProProSerGlyProGlyArgPheProValProSer 63
1136 TCCCAAGTACGCTGCTTCGCGTGAAGCATTCCTGACCGGCTTCGCTG 1087
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
64 ProGlyHisArg.....SerGlyProProGluProGluArgAsnArg... 77
1086 ATCGTCGCGCGCAGGTTCTGACGCGGTGTTCAGCTCGGTAGCGGTGCG 1037
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
78 .....MetPro.....ProProA 82
1036 GTCCCATTTTTCCTGGACACCTGCTGTAGC.....CCTCCGAACGCTAC 993
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
82 rgProAspValGlySerLysProAspSerIleProProValProSer 98
992 CGCCCGACGCGCTCGCA...GCTTGTGTCAGGACTGCTTCCCTCGTCGA 946
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
99 ThrProArgProIleGlnSerSerLeuHisAsnArgGlySerProProva 115
945 AGGAGGAATGAATGACGTGACATTTCCCTGGATTGCGCTTGGCGCGG. 897
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
115 lProgly.....GlyProArgGlnProSerP 124
896 .....CCTCATACCGCGCAATTCACACTGCTGCTCTGTCA 861
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
124 roGlyProThrProProProPheProGlyAsnArgGlyThrAlaLeuGly 140
860 TGTTTTTCCTCGCTTCTTTTCGATTAGCGGTCAGAAAGCCATTTTCG 811
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
141 GlyGlySerIleArgGlnSerProLeuSerSerSer...SerProPheSer 156
810 AGGACAGCGCCTGCT.....GCTGC 791
157 AsnArgProProLeuProProThrProSerArgAlaLeuAspAspLysPr 173
790 TCTCGTCGCGCCTCGAGTATTGGACGCGCGCTGACGAATATTCGTGCA 741
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
173 oProProProProProProValGlyAsnArgPro.....SerIleH 187
740 GATCTCGTCGAGTTCCTGCTCTGCTATTGCTGCTTCTTGAAGCGCA 691
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
187 isArgGluAlaValPro..... 192
690 CCACCGCGCTGGCGCGCTCCCGCGCGCGCGCTGACCTGCGCCTGC 641
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
193 ProProProProGlnAsnAsnLysProPro..... 202
640 AACGAACCTGCGCTGACCTCCACCTGGTCGATCTGGGTTTTCAGGTGCGC 591
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
203 .....ValProSerThrPro.....A 208
590 GGAGATCGCT.....CGAAATACCTGCTCCTCGCCGCGCGAGGTAGCGG 547
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
208 rgProSerAlaProHisArgProHisLeuArgPro...Pro..... 220
546 CATCGGTCTTCATCTCTGCCATGTGCTGACTACTTCTCTCTTTACCTTC 497
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
221 .....ProPro 222
496 TCGCAAAATGTTGGCAAGTCTTCGCGCGGTTGGCGGGAAGTCTGTT 447
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
223 Ser.....ArgProGlyProPro..... 228
446 GTCAATTACGGAGCTCACAGTCGCTCTTCTGTCCTCCAGTCGT..... 404
229 .....ProLeuProProSerSerSerGlyA 237
403 .....CCTGCTGCTTCTTTCACGCTCCTCGCGGA 374
237 snAspGluThrProArgLeuProGlnArgAsnLeu...SerLeuSerSer 252
373 CGGTGTCGCGCGCGACACCGCGGCTGGTGGACCGCGCGGATTGCGGAA 324
253 SerThrProProLeuProSerProGly.....Ar 262
323 CCTTGGCCATCGCTCCCGGACCCACCGGCGCGCCACCGTCACCGA 274
262 gSerGlyPro...LeuProProValProSerGluArgProProProp 278
273 CGATCCGCGCAACAGCGCGCGCATCA.....CCGAGGGGGCAACCG 233
278 roValArgAspProProGlyArgSerGlyProLeuProProProPro 294
232 GCTTTTCATCAGCTGACATCAGCGCGCTGCGGTCAACGACCCACCT 183
295 ValSerArgAsnGlySerThrSerArgAlaLeuProAlaThr...ProGln 310
182 GCGCA.....GGTAGCGACTCCCGCGCGCAGCGCGCGCGCGCG 142
311 LeuProSerArgSerGlyValAspSerProArgSerGlyProArgProPr 327
141 GCTGGGCGCTGATCCACCGCAGCGGATGTTGACAGCGGACTGGTGC 92
327 oLeuProProAspArgPro..... 333
91 CGACGAGCCCATCTGCGCGGCTTCTGCTGCTGGGTGCGCGCGCGCG 42
334 .....SerAlaGlyAlaProProPro 340
41 GTGCGCGCCACC 30
341 ProProProSer 344
seq_name: pir2:T02345
seq documentation_block:
hypothetical protein KIAA0324 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 05-Nov-1999
C:Accession: T02345
R:Ricke, D.O.; Bruce, D.; Mundt, M.; Doggett, N.; Munk, C.; Saunders, E.; Robinson, D.
re, J.; White, S.; Ueng, S.; Tatum, O.; Campbell, C.; Fawcett, J.; Deaven, L.
submitted to the EMBL Data Library, March 1998
A:Description: Sequencing of human chromosome 16p13.3.
A:Reference number: Z14664
A:Accession: T02345
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1791 <RIC>
A:Cross-references: EMBL:AC004493; NID:g2996648; PIDN:AC08453.1; PID:g2996650
C:Genetics:
A:Map position: 16
A:Introns: 1610/2; 1706/2
A>Note: KIAA0324
```

alignment\_scores:  
Quality: 179.50 Length: 493  
Ratio: 0.820 Gaps: 25  
Percent Similarity: 44.422 Percent Identity: 26.572

alignment\_block:  
US-09-462-480-1/rev x T02345 ..



Align seg 1/1 to: T02345 from: 1 to: 1791

```
1271 TCGGCGCGCATGACAACTCTCAGAGTCGGCTCAAACTGATAAACCAGAG 1222
:::|||||:::
1269 AlaAlaAlaMetAsnLeuAlaSerAlaArgThrProAlaIleProThrAl 1285
:::|||||:::
1221 AAAGGCGAGACCGAGCGGAAGTGAAGTCCGCCGATCCCGTTCGCT 1172
:::|||||:::
1285 aValAsnLeuAlaAspSerArgThrProAlaAlaAlaAlaMetAsnL 1302
:::|||||:::
1171 AT...TCTACGGGAAGTCCGGTTCCTATCGGAACATCCAGTCAGGT 1125
||| ||||| ::||| ||
1302 euAlaSerProArgThrAlaValAla.....ProSerAlaVal 1314
:::|||||:::
1124 TGCCTTCGGTCAAGCCATTGCTCAGCGGTTCGCTGATCGTCGCCGCC 1075
:::|||||:::
1314 l.....AsnLeuAlaAsp...ProArgT 1321
1074 AGTTTCTCAGCGCGTGTTCAGTCTGGTAGCGGTGGCTCCCAVTTTGT 1025
:::|||||:::
1321 hrProThrAlaProAlaValAsnLeuAlaGlyAlaArgThrProAlaAla 1337
1024 CTGGACACCTGGTACGCTCGAACCCTACCGCCCGCCAGCGCTGCGA 975
||| ::||| ::||| ||| ::||| |||
1338 LeuAlaAlaLeuSerLeuThrGlySerGlyThr.ProProThrAlaAla 1354
:::|||||:::
974 GCTTGGTCAGGACTGCTTCCCTCGTCAAGGAGGGAATGAATGACGCTG 925
:::|||||:::
1354 sn.....TyrProSerSerSerArgThrProGln..... 1363
924 ACATTTCCCTGGATTGCGCTTGCCTGCCCGCGCTCGATA.....CCGCGAA 881
||| ||||| ::||| ::|
1364 .....AlaProAlaSerAlaAsnLeuValGlyProArgSe 1375
880 ATTCCACTGCTCTCTGATGATTTTTCCTCGGTTCCTTCGTATTACG 831
||| ::|
1375 rAlaHisAlaThrAla.....P 1381
830 GGGTCAGAAGCCATTTCGAGGACAGCGCTGCTGCTGCTCTCGTCGG 781
||| ::| ::||| ||||| ||| |||||
1381 roValAsnIleAlaGlySerArgThrAlaAlaAlaLeuAlaProAlaSer 1397
780 CCTCGAGTATGACGGCGGCT.....GACGAATATTCGTCGAGATC 737
||| ||| ::|
1398 LeuThrSerAlaArgMetAlaProAlaLeuSerGlyAlaAsnLeuThrSe 1414
736 TCGTCGAGTCTCTGCTCTGCTTATGCTGCTTCTTGA....ACGCGAC 690
||| ||||| ||||| ||||| ::|
1414 rProArgValProLeuSerAlaIleTyrGluArgValSerGlyArgThrSerP 1431
689 CACCG.....CGGCTGGCGCGCGCTCC 667
||| |||||
1431 roProLeuLeuAspArgAlaArgSerArgThrProProSerAlaProSer 1447
666 CGCGCGCG.....CGCGCCTGCGCTGCAACGAACTGCC 629
||| ||||| ::||| ::||| ::|||
1448 GlnSerArgMetThrSerGluArgAlaProSerProSerSerArgMetG 1464
628 GTCGACTCCA..... 619
:::|||||
1464 yGlnAlaProSerGlnSerLeuLeuProProAlaGlnAspGlnProArgS 1481
618 ..CCTGGTCTGCTGGTGTTCAGTTCGGCGGAGATCCGCTCGAAATTAC 571
||| ||||| ::|||
1481 erProValProSerAlaPheSerAspGlnSerArgCysLeuIleAlaGln 1497
570 CTGCTCTCTGCGGAGGTAGCGGATCGGTCTTCATCTCTGCCATGCTG 521
||| ::||| ::||| ::|||
1498 ThrThrProValAlaGlySerGlnSerLeuSerSerGlyAlaValAlaTh 1514
520 GACTACTTCTCTCT..... 505
||| |||||
1514 rThrThrSerSerAlaGlyAspHisAsnGlyMetLeuSerValProAlaP 1531
```

```
504 ..TACCTTCTCGCCAAATGTTGCAAGTCTTCGGCCCGGGTGGCCGG 457
:::|||||:::
1531 roGlyValProHisSerAspValGlyGluPro.ProAlaSer..... 1544
456 GAAGTCTCTTGTCAATTACGGGAGCTCACAG..... 426
|||||:::
1545 .....ThrGlyAlaGlnProSerAlaLeuAlaLaL 1556
425 .....TCGTCC 420
1556 euGlnProAlaLysGluArgArgSerSerSerSerSerSerSerSer 1572
419 TCTTCGTCCAGTCTGCTCGCTCTTCTTTCACGCTCTGCGCGAGCGG 370
|||||:::
1573 SerSerSerSerSerSerSerSerSerSerSerSerSerSerSerGl 1589
369 TGCCGGCGCGACACCGCGGCTGGTGAGCCCGCGGATTCGGAACCTT 320
|:::|||||::|
1589 ySerSerSerSer.Asp.....SerGlu 1596
319 GGCCCATCGCTCCCGACCCACCGGAG.....CGCGCCACCCGTC 279
||| ||||| |||||
1597 GlySerSerLeuProValGlnProGluValAlaLeuLysArgValProSe 1613
278 ACCGAGATCCGGCAACAGCCCGCGCATCACCGAGGGGCAACCGGCTT 229
||||| ||||| ::|||
1613 rProThr..ProAlaProLysGluAlaValArgGluGlyArg..... 1626
228 TTCGATCAGCTGAGACATCAGCGGCGTGGGTCAACGACCCACCTCGCG 179
||| |||||
1627 .....ProProGluP 1630
178 CA.....GGTAGCGACTCCGCGCGCAGCAGG 153
||
1630 roThrProAlaLysArgLysArgArgSerSerSerSerSerSerSer 1646
152 CCGCGCGCGCTGGGCTGATCACCGACGCGCGGATGTTGACAG 103
::: ||| ::|
1647 SerSerSerSerSerSerSerSerSerSerSerSerSerSerSe 1663
102 CGGACTGGTGGCGAGCAGCGCCATCTGCGGCTCTCTCGTGGCTGGT 53
|::: ||| :::|||||:::
1663 rSerSerSerSerSerSerSerSerSerSerSerSerSerSers 1680
52 TGCCGCGCGCGTGGCGGCC 33
||| |||||
1680 erProSerProAlaLysPro 1686
```

seq\_name: pir2:B40505

seq\_documentation\_block:

hypothetical protein - suid herpesvirus 1 (strain Indiana-Funkhuser or Becker)  
C/Species: suid herpesvirus 1  
C/Date: 10-Apr-1992 #sequence\_revision 10-Apr-1992 #text\_change 02-Sep-2000  
C/Accession: B40505  
R/Cheung, A.K.

J. Virol. 65, 5260-5271, 1991

A/Title: Cloning of the latency gene and the early protein 0 gene of pseudorabies vi  
A/Reference number: A40505; MUID:91374576

A/Accession: B40505

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-1958 <CHE>

A/Cross-references: GB:M57505; NID:g334066; PIDN:AAA47468.1; PID:g334068

C/Superfamily: pseudorabies virus 1 nuclear antigen

alignment\_scores:

Quality: 178.50 Length: 508

Ratio: 0.776 Gaps: 32

Percent Similarity: 45.276 Percent Identity: 28.346

## alignment\_block:

US-09-462-480-1 x B40505

Align seg 1/1 to: B40505 from: 1 to: 1958

```
31 GTGCGGCGCACCGCC...GGCGGCAACCCA..... 57
|||||
1053 ValGlyGlyGluGlyArgGlyGlyGlyProArgValGlyLeuAlaG1 1069
58 .....GCCGACGAGGAAGCGCGCAGATGGCTGTCTGCGCACCACTC 100
|||||
1069 yArgAspAlaAlaGluAlaAlaValAlaGlyArgGlyValLeuGlyHisGlyP 1086
|||||
101 CG...CTGTCCGAACCATCG.....CTGGCTGGTGA..... 129
||
1086 roGluArgAlaProGluProValValLeuGlyGlyGlyGlyGlyGly 1102
130 .....TCAGGCCCC...AGCGCGGG 146
1103 GlyGlyHisGluArgGlySerGlyValArgSerGlyProGluSerGluG1 1119
147 CGCGGCGCTG..... 156
1119 yAlaAlaLeuAlaProGlyProProValLeuPheValValAlaValAlav 1136
157 .....CTGCGCGCGGAG.....TCG 171
|||
1136 aAlaValProAlaGluGlyArgAlaGlyGluProLeuValLeuAla 1152
172 CTACCTCGCGCGAGGTGGTGTGACCGCGCGCTGTGATCTCTCAGCT 221
|||||
1153 ValProGlyAlaAlaGly.....ProGlyArgAlaAlaLe 1164
222 GATCGAAAGCGGTGCCCCCTCGTGTATGCGCGCGCTGTCTGCGGAT 271
|||
1164 uLeuLeuAlaProLeuGlyArgTrpValArg...AlaGlyGlyGlyGly 1180
272 CGTGGTGCAGCGGTGCGCGCTCGGTGGTTCGGGAGCG..... 312
|||||
1180 laGlyValAlaGlyGlyAlaGlyGluAlaGlyLeuGlyAlaGlyAlaGly 1196
313 .....ATGGCGCAGGTTCCGATCCCGCGGCTCCAC..CAGCCCGGTC 355
|||
1197 LeuGlyAlaGlyAlaGlyLeuGlyAlaGlyAlaGlyGlyProGlyAl 1213
356 TGGTCGCGCGCGCACCGCTCGCGCAGGAGCGTGAAGAAGACGA..... 398
|||||
1213 aGlyGluAlaGly.....GlyGlyAlaArgArgArgArgArgA 1227
399 .....CGAGGACGACTGGGA 413
1227 rgTrpAspAlaGlyLeuLeuGlyProGluArgGlyGluAlaGly 1243
414 CGA.....AGAGGACGACTGGTGTGAGCTCCGCTAA 442
|||
1244 ArgGlyLeuArgGlyProGlyProArgGlyGlyLeuGlyGluProGly.. 1259
443 TGACAACAGACTCCCGGCGCACCGCGCGGAAGACTTGCCCAACATTGTG 492
|||||
1260 .....ProGlyHisValGlyArg.....G 1266
493 CGGAGGAAGGTAA..... 506
1266 lyGluGluGlyArgGlyValGlyProGlyGlyLeuAlaGlyAlaGlyPro 1282
507 .....GAGAGAAGTAGTCCAGCATGGCAGATGAA 538
|||||
1283 ValHisAlaValAlaHisGluArgHisGlyAla..GlyAspGluGly 1298
539 GACCGATCCGCTACCTCGCGGCGAGGAGGTAAATTCGAGCGGATCT 588
|||||
1299 AspArgValArgGlyLeuLeuProLeuGlyArgAlaGlyProGlyAsp.. 1314
```

```
589 CCGGCGACCTGAAACCCAGATCGACCAAGGTGGAGTCTGACGCGCAGGTTGG 638
|||
1315 ArgValAlaGluArgGluGlnArgGlyGlyHisLeu.....LeuGluA 1329
639 TTGACAGGCCAGTGGCGCGCGC.....GGCGGAGCGCGCGCCAGCG 682
|||||
1329 laGlyGlyProGluGlyArgGlyAlaGlyGlyArgGlyGlnProGlu 1345
683 CGCGGTGGT.....GCGTTCGAAGACGACCAATAGCAAGACAGCAGG 726
|||||
1346 ArgAlaGlyGlnGlnAlaLeuGluAspAlaAlaGlyGlnAspAlaG1 1362
727 AACTCGACGAGATCTCGAGAAATATTCGTAGCGCGGCTCCCAATACTCG 776
|||||
1362 yValArgGln...LeuAlaGlyHisAlaAlaGlyLeuArg.....G 1375
777 AGGCGCGGAGGAGGACGACGAGCGCGTGTCTCGCAATGGGTTCTG 826
|||||
1375 lyGlyGluGlyGlyAlaAspAlaGlyAlaGluGlyLeuAspGlyArgLeu 1391
827 ACCCGCTAATACGAAAGAAAGCGGAGCAAAACATGACAGACGACGAGTG 876
1392 ProGly..... 1393
877 GAATTCGCGGCTATCGAGCGCGCGCAATCCAGGAAATGTCA 926
|||||
1394 .....AlaGlyValArgGlyAlaAlaArgValGlyHisValGlyValG 1408
927 GTCCATTCATTCCTCTTACGAGGGAAGCAGTCCCTGACCAAGCTC 976
|||||
1408 lyProAlaGluAlaLeuGlnAspGluGly.....LeuLeuGlyAlaAla 1422
977 GCAGCGCGCTGGCGGTAGCGGTTCGAGCGGTACCC..... 1013
1423 ValAlaAlaAlaHisGlyHisGlyAla..HisArgValArgGlnGlyProG 1439
1014 ..AGGTTCTCCAGCAAAATGGAGCGCGCTACCGAGCTGAACAACG 1061
|||||
1439 luArgValLeuGlyGlyHisGlyValProAspVal..... 1450
1062 CGTGCAGAACCTGGCGGAGGAGTACCGAAGCGCGGTACGCAATGGCT 1111
|||
1451 .....ArgGlnArgArgGlyHisAlaAlaGlyAsp..... 1460
1112 TCGACCGAAGCAACGCTCACTGGGATGTCGATAGGCAACGCGCGAGTT 1161
|||||
1461 .....GluGlyAlaValAlaValGly.....ArgValAspProAlaL 1473
1162 CGCGTAGAATAGCAAAACACGGGATCGGCGGAGTTCGACCTTCCGTC... 1208
|||||
1473 euAlaGluLeuValGluAlaLeuValGlyGlyLeuHisProArgValLeu 1489
1209 .....GGTCTCGCC 1217
1490 GlnProGlyHisGlyLeuAla 1496
```

seq\_name: pir2:T08179

## seq\_documentation\_block:

LRC5 protein - Chlamydomonas reinhardtii

C:Species: Chlamydomonas reinhardtii

C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 11-Jun-1999

C:Accession: T08179

R:Glockner, G.; Beck, C.F.

submitted to the EMBL Data Library, October 1996

A:Description: Molecular characterization of a gene (LRC5) involved in blue light sig

A:Reference number: Z16399

A:Accession: T08179

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-640 &lt;GLO&gt;

A:Cross-references: EMBL:U73817; NID:g1644369; PID:g1644370

C:Genetics:



A:Title: cDNA cloning and chromosomal mapping of the mouse type VII collagen gene (Col7A)  
A:Reference number: A45748; MUID:93315168  
A:Accession: A45748  
A>Status: preliminary  
A:Molecule type: nucleic acid  
A:Residues: 1-920 <L11>  
A:Cross-references: GB:S63654; NID:g386656; PIDN:AAB27492.1; PID:g386657  
A:Experimental source: epidermal keratinocyte  
A>Note: sequence extracted from NCBI backbone (NCBIN:135000, NCBIP:135001)  
C:Superfamily: unassigned collagens

## alignment\_scores:

Quality:	174.00	Length:	421
Ratio:	0.911	Gaps:	24
Percent Similarity:	45.368	Percent Identity:	26.603

## alignment\_block:

US-09-462-480-1 x A45748 ..

Align seg 1/1 to: A45748 from: 1 to: 920

```
27 CCAGTGGCGGCGCCGCGCGCCGCAACCCAGCGAGCGAGCGCGC 76
||||| :|||:|||||:|||||:|||||:|||||:|||||:
292 ProGlyProGlnGlyProAlaGlyArgProGlyGluLysGluLysG 308
77 AGAT.....:|||||:|||||:|||||:|||||:
308 yAspCysGluAspGlyGlyProGlyLeuProGlyGlnProGlyPro. 324
91 GGCACAGTCGCGTGTGCAACCATCCGCTGGCTGGTATCAGCCCGC 140
||| :|||:|||||:|||||:|||||:|||||:
325 GlyGluProGlyLeuArgGlyAlaPro.....GlyMetThrGlyPro 339
141 CGCGGCGCGCGCGCTGCTCGCGCGGAGTCGCTACCTGCGCGAGTGG 190
:|||||:|||||:|||||:|||||:|||||:
339 sGlyAspArgGlyLeuThrGlyThrProGlyGluProGlyValLysG 356
191 CGTTGACCGCGCGCGCTGATGTCTCAGCTGATCGAAAGCCGTTGCC 240
:|||||:|||||:|||||:|||||:
356 LuArgGlyHis.....ProGlyProValGly 364
241 CCCTCGGTGATCGCGCGCGCTGTCGCGGATCGCTGCGTACGCGCGC 290
|||||:|||||:|||||:|||||:|||||:
365 ProGlnGlyLeuProGlyAlaAlaGlyHisProGlyValGluGlyPro 381
291 CGCTCCGCGTGGTCCG.....:|||||:|||||:
381 uGlyProProGlyProThrGlyArgGlyGluLysGlyGluProGlyA 398
307 .....:|||||:|||||:|||||:
398 rgProGlyAspProAlaValGlyProGlyGlyAlaGlyAlaLysGly 414
322 GGTTCGCAATCCGC.....GGCTCCACC.. 345
:|||||:|||||:|||||:
415 LysGlyGluAlaGlyLeuProGlyProArgGlyAlaSerGlySerLys 431
346 .....AGCCCGGTCTGTCGCGCGCACCCGCTCGCGCAGGAGC 385
:|||||:|||||:|||||:
431 yGluGlnGlyAlaProGlyLeuAlaLeuProGlyAspProGlyProLys 448
386 GTGAAGAAGACGACGAGG.....ACGACTGGGACGAAGAGGAC 423
:|||||:|||||:|||||:
448 LyAspProGlyAsp.ArgGlyProIleGlyLeuThrGlyArgAlaGly 464
424 GACTGGTGGCTCCGTAATGACAACAGACTTCCGCGCCACCGGCGCG 473
|||||:|||||:|||||:
464 oThrGlyAsp.....SerGlyProProGlyGluL 474
474 AAGACTTGCCAACATTTTGGCGAGGAAGTTAAAGAGAG.....AAAGTA 517
:|||||:|||||:|||||:
474 ySgGlyGluPro.....GlyArgProGlySerProGlyProVal 486
```

```
518 GTCCAGCATGCGCAGAGATGAAGACCGATGCGGCTACCCCTCGGCGCAGGAG 567
:|||||:|||||:|||||:|||||:
487 GlyProArgGlyArgAspGlyGlu.....AlaGlyL 497
568 CAGGTAATTTTCAGCGGATCTCCGCGCGACCTGAAACCCAGATCGACGAG 617
:|||||:|||||:|||||:|||||:
497 u.....LysGlyAspGluGlyIleProGlyGluProGlyLeuProG 511
618 GTGGAGTCGCGCGCAGGTTTCGTT.....GCAGGGGCA 649
||| :|||:|||||:|||||:
511 LyLysAlaGlyGluArgGlyLeuArgGlyAlaProGlyProArgGlyPro 527
650 GTGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGCGCTTCC 699
|||||:|||||:|||||:|||||:
528 ValGlyGluLysGlyAspGlnGlyAspProGlyGluAspGly..... 541
700 AAGAAGCAGCAATGAAGCAGAGCAGCAATTCGACGAGATCTCGACGAAT 749
|||||:|||||:|||||:|||||:
542 .ArgAsnGlySerProGlySerSerGlyProLysGlyAspArgGlyGluP 558
750 ATTGTCAGGCGCGCGCTCAATCTCGAGGCGCGCGAGGAGCAGCAGCA 799
:|||||:|||||:|||||:
558 roGlyProGlyProPro.....GlyArgLeuValAspAla 570
800 GGCCTGCTCTCGCAAAATGGGCTTCTGACCGCGCTAAATACGAAAGAAACG 849
||| :|||||:|||||:|||||:
571 Gly.IleGluSerArgAspLysGlyGluProGlyGlnGluGlyProArg 587
850 GAGCAAAACATGACAGCAGCAGCAGTGAATTTCCGCGGTTATCGAGGC... 897
||| :|||||:|||||:|||||:
587 lyProLysGlyAspProGlyProGlyValSer.GlyGluArgGlyI 603
898 .....CGCGCAAGCGCAATCCAGGGAATGTACCTCCATTTCAT 937
||| :|||||:|||||:|||||:
603 eAspGlyLeuArgGlyProProGlyPro.....GlnGlyA 615
938 TCCCTCTTTCAGCAGGGAAGCA.....GTCCCTGACCAAGCTCGCAGC 981
||| :|||||:|||||:|||||:
615 spProGlyValArgGlyProAlaGlyAspLysGlyAspArgGlyProPro 631
982 GGCCTG...GGCGGTAGCGGTTCCGAGGCGTACCAGGTTCTCCAGCAAA 1028
|||||:|||||:|||||:|||||:
632 GlyLeuAspGlyArgSerGlyLeuAspGlyLysProGlyAlaProGlyPr 648
1029 AATGGGACGCCAGCTACCGAGCTGAACAACGCGCTGCAGAACTGGCG 1078
||| :|||||:|||||:|||||:
648 oProGlyLeuHisGly.....AlaSerGlyLysAlaGlyAspProGlyA 663
1079 CGGAC 1083
|||
663 rgAsp 664
```

OM of: US-09-462-480-1 to: SwissProt\_40.\* out\_format : pfs  
Date: Jul 22, 2002 1:41 AM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:  
-MODE=frame+nt2p.model -DEV=xlh  
-O=/cgn2.1/USPto\_spool/US09462480/runat\_18072002\_164420\_19600/app\_query.fasta\_1.2850  
-DB=Swissprot\_40 -QFWT=fastan -SUFFIX=rsp -GAPOP=12.000  
-CAPEXT=4.000 -MINMATCH=0.100 -LOOPECL=0.000 -LOOPEXT=0.000  
-GGAPOP=4.500 -GGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFWT=pfs  
-NORM=ext -HEARSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09462480 -CGNI\_1\_29 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

Search information block:  
Query: US-09-462-480-1  
Query length: 1277  
Database: Swissprot\_40.\*  
Database sequences: 105224  
Database length: 38719550  
Search time (sec): 49.760000

score_list:	Sequence	Strd Orig	zscore	EScore	Len	! Documentation
	SwissProt_40:CF10_MYCTU	481.00	471.11	1.5e-18	99	! 069739 mycobacterium tubercul
	SwissProt_40:ES66_MYCTU	474.00	464.91	3.6e-18	94	! 057165 mycobacterium tubercul
	SwissProt_40:YHL1_EBV	212.50	204.07	0.0002	660	! P03181 Epstein-Barr virus (str
	SwissProt_40:WAP1_HUMAN	189.50	184.45	0.0028	503	! O43516 Homo sapiens (human)
	SwissProt_40:ES66_MYCLE	186.00	193.64	0.0046	95	! O50206 mycobacterium leprae
	SwissProt_40:CF10_MYCLE	182.00	189.56	0.0074	99	! O32084 mycobacterium leprae
	SwissProt_40:VNUA_PVKA	172.50	159.18	0.0208	1733	! P33485 pseudorabies virus (
	SwissProt_40:DRPL_RAT	170.50	160.15	0.0268	1183	! P54258 rattus norvegicus (rat
	SwissProt_40:IRIS2_HUMAN	169.50	158.37	0.0301	1324	! Q9Y4H2 Homo sapiens (human)
	SwissProt_40:T2D3_HUMAN	168.50	158.93	0.0343	1083	! O00268 Homo sapiens (human)
	SwissProt_40:DRPL_HUMAN	168.00	154.96	0.0521	1185	! P54259 Homo sapiens (human)
	SwissProt_40:CAFE_RIFPA	164.50	155.56	0.0557	1027	! P30754 Riftia pachyptila (b
	SwissProt_40:CA21_BOVIN	164.50	153.44	0.0551	1364	! P02465 Bos taurus (bovine)
	SwissProt_40:MY15_HUMAN	163.50	145.37	0.0598	3530	! Q9UKN7 Homo sapiens (human)
	SwissProt_40:CA17_HUMAN	163.00	146.26	0.0640	2944	! Q02388 Homo sapiens (human)
	SwissProt_40:EXLP_TOBAC	162.50	160.27	0.0734	426	! Q03211 Nicotiana tabacum (c
	SwissProt_40:VRP1_YEAST	161.00	153.98	0.0857	817	! P37370 Saccharomyces cerevis
	SwissProt_40:CA21_RANCA	160.50	149.72	0.0893	1355	! O42350 Rana catesbeiana (bu
	SwissProt_40:IE18_PVIF	160.50	149.16	0.0890	1461	! P11675 pseudorabies virus (
	SwissProt_40:CA21_MOUSE	159.50	148.69	0.1007	1372	! Q01149 Mus musculus (mouse)
	SwissProt_40:D4DR_HUMAN	159.00	156.29	0.1116	467	! P21917 Homo sapiens (human)
	SwissProt_40:IE18_PVKA	158.50	147.35	0.1134	1446	! P33479 pseudorabies virus (
	SwissProt_40:CA25_HUMAN	158.50	147.10	0.1132	1496	! P05997 Homo sapiens (human)
	SwissProt_40:GP1_CHURE	157.50	153.58	0.1328	555	! Q9FQ6 Chlamydomonas reinhard
	SwissProt_40:ICP0_HSV2H	157.50	150.61	0.1308	825	! P28284 herpes simplex virus
	SwissProt_40:IE18_PVKA	156.50	145.47	0.1443	1446	! P33479 pseudorabies virus (
	SwissProt_40:S3A2_MOUSE	156.00	153.33	0.1602	475	! Q62203 Mus musculus (mouse)
	SwissProt_40:CA21_RAT	156.00	145.39	0.1536	1372	! P02466 rattus norvegicus (rat
	SwissProt_40:CA11_CHICK	155.00	144.96	0.1533	1453	! P02457 Gallus gallus (chick
	SwissProt_40:S3A2_HUMAN	155.00	152.57	0.1809	464	! Q15428 Homo sapiens (human)
	SwissProt_40:EXTN_TOBAC	155.00	150.40	0.1789	620	! P13983 Nicotiana tabacum (c
	SwissProt_40:CA12_MOUSE	155.00	143.99	0.1729	1459	! P28481 Mus musculus (mouse)
	SwissProt_40:SES_RAT	153.00	146.37	0.2252	825	! Q63003 rattus norvegicus (rat)
	SwissProt_40:CA21_HUMAN	153.00	142.60	0.2208	1366	! P08123 Homo sapiens (human)
	SwissProt_40:IE18_PVIF	153.00	142.09	0.2202	1461	! P11675 pseudorabies virus (
	SwissProt_40:M3K1_RAT	153.00	141.93	0.2200	1493	! Q62925 rattus norvegicus (rat
	SwissProt_40:GSRI_HUMAN	152.50	141.38	0.2336	1509	! Q9NZM4 Homo sapiens (human)
	SwissProt_40:SPD1_NEPL	152.00	146.18	0.2551	747	! P19837 Nephila clavipes (ort
	SwissProt_40:SPD1_ONCMY	152.00	141.71	0.2492	1356	! Q93484 oncorhynchus mykiss
	SwissProt_40:ICP0_HSV11	151.00	144.96	0.2874	775	! P08393 herpes simplex virus

SwissProt\_40:CA11\_MOUSE + 151.00 140.25 0.2804 1453 ! P11087 mus musculus (mo  
SwissProt\_40:ICP0\_HSV2H + 150.50 144.02 0.3046 825 ! P28284 herpes simplex vi  
SwissProt\_40:CA13\_HUMAN + 150.50 139.71 0.2977 1466 ! P02461 homo sapiens (hu  
SwissProt\_40:CA1B\_MOUSE + 150.50 138.16 0.2953 1804 ! Q61245 mus musculus (mo  
SwissProt\_40:CA13\_BOVIN + 150.00 141.75 0.3205 1049 ! P04258 bos taurus (bovi

seq\_name: SwissProt\_40:CF10\_MYCTU

seq\_documentation\_block:

ID CF10\_MYCTU STANDARD; PRT; 99 AA.  
AC 069739;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DE 16-OCT-2001 (Rel. 40, Last annotation update)  
DE 10 kDa culture filtrate antigen cfp10.  
GN CFP10 OR LHP OR RV3874 OR MT3988 OR MTV027.09.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-15.  
RC STRAIN=H37RV;  
RX MEDLINE=99061212; PubMed=9846755;  
RA Berthet F.X., Rasmussen P.B., Rosenkrands I., Andersen P.,  
RA Gicquel B.;  
RT "A Mycobacterium tuberculosis operon encoding ESAT-6 and a novel  
RT low-molecular-mass culture filtrate protein (CFP-10).";  
RL Microbiology 144:3195-3203(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H37RV;  
RX MEDLINE=98295987; PubMed=9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekaita F.,  
Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
Davies K., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
Hornby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
RT complete genome sequence.";  
RL Nature 393:537-544(1998).  
[3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CDC 1551 / Oshkosh;  
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
Bishai W.;  
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
RT laboratory strains.";  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
CC -! SIMILARITY: BELONGS TO THE ESAT6 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AF004671; AAC83445.1; -  
CC EMBL; AL022120; CAAL17966.1; -  
CC EMBL; AE007190; AAK48356.1; -  
CC TIGR; MT3988; -  
CC TubercuList; RV3874; -  
CC Complete proteome.  
CC INIT\_MET 0 0  
CC SEQUENCE 99 AA; 10663 MW; EBCAB6A996C5489D CRC64;

alignment\_scores:  
Quality: 481.00 Length: 99  
Ratio: 4.859 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 98.990

alignment\_block:  
US-09-462-480-1 x CF10\_MYCTU ..

Align seg 1/1 to: CF10\_MYCTU from: 1 to: 99

528 GCAGAGATGACGACGATGCGCTACCTCGGCGGAGGAGCGAGTAATT 577  
|||||  
1 AlagluMetLysThrAspAlaAlaThrLeuAlaGlnGluAlaGlyAsnPh 17  
|||||  
578 CGAGCGGATCCGCGGACCTGAAACCCAGATCGACCGAGTGGAGTCGA 627  
|||||  
17 eGluArgIleSerGlyAspLeuLysThrGlnIleAspGlnValGluSer 34  
|||||  
628 CGGCGAGTTCGTCAGGCGCAGTGGCGCGCGCGGCGGAGCGCGCGCC 677  
|||||  
34 hrAlaGlySerLeuGlnGlyGlnTrpArgGlyAlaAlaGlyThrAlaAla 50  
|||||  
678 CAGCGCGGTCGTCGCTCCAGAGACGCCCAATAGCAGAGCAGGA 727  
|||||  
51 GlnAlaAlaValValArgPheGlnGlnAlaAlaAsnLysGlnLysGln 67  
|||||  
728 ACTCGAGGATCGCAGGATATCTGTCAGCGCGGCTCAATACCTCGA 777  
|||||  
67 uLeuAspGlnIleSerThrAsnIleArgGlnAlaGlyValGlnTrpSer 84  
|||||  
778 GGGCGCGGAGGAGCAGCAGCAGCGCTGCTCGCAATGGGCTTC 824  
|||||  
84 rgAlaAspGluGlnGlnGlnAlaLeuSerSerGlnMetGlyPhe 99  
|||||

seq\_name: SwissProt\_40:ESA6\_MYCTU

seq\_documentation\_block:  
ID ESA6\_MYCTU STANDARD; PRT; 94 AA.  
AC Q57165; O84901;  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE 16-OCT-2001 (Rel. 40, Last annotation update)  
DE 6 kDa early secretory antigenic target (ESAT-6).  
GN ESAT6 OR RV3875 OR RV3989 OR RV027.10.  
OS Mycobacterium tuberculosis, and  
OS Mycobacterium bovis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773, 1765;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES=M.tuberculosis; STRAIN=H37RV;  
RX MEDLINE=95204931; PubMed=7897219;  
RA Andersen P., Andersen A.B., Sorensen A.L., Nagai S.;  
RT "Recall of long-lived immunity to Mycobacterium tuberculosis infection  
in mice.";  
RL J. Immunol. 154:3359-3372(1995).  
RN [2]  
RP SEQUENCE FROM N.A., SEQUENCE OF 1-10, AND CHARACTERIZATION.  
RC SPECIES=M.tuberculosis; STRAIN=ERDMAN;  
RX MEDLINE=95247251; PubMed=7729876;  
RA Soerensen A.L., Nagai S., Houen G., Anderson P., Anderson A.B.;  
RT "Purification and characterization of a low-molecular-mass T-cell  
antigen secreted by Mycobacterium tuberculosis.";  
RL Infect. Immun. 63:1710-1717(1995).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC SPECIES=M.tuberculosis; STRAIN=H37RV;  
RX MEDLINE=98295987; PubMed=9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,

RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
complete genome sequence.";  
RL Nature 393:537-544(1998).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;  
RA Fietischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
RA Bishai W.;  
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
laboratory strains.";  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A.  
RC SPECIES=M.bovis;  
RX MEDLINE=96200095; PubMed=8631702;  
RA Mahairas G.G., Sabo P.J., Hickey M.J., Singh D.C., Stover C.K.;  
RT "Molecular analysis of genetic differences between Mycobacterium bovis  
BCG and virulent M. bovis.";  
RL J. Bacteriol. 178:1274-1282(1996).  
RN [6]  
RP SEQUENCE OF 1-70 FROM N.A.  
RC SPECIES=M.tuberculosis; STRAIN=H37RV;  
RX MEDLINE=99061212; PubMed=9846755;  
RA Berthet F.-X., Rasmussen P.B., Rosenkrands I., Andersen P.,  
RA Gicquel B.;  
RT "A Mycobacterium tuberculosis operon encoding ESAT-6 and a novel  
low-molecular-mass culture filtrate protein (CFP-10).";  
RL Microbiology 144:3195-3203(1998).  
CC -!- FUNCTION: NOT KNOWN. ELICITS HIGH LEVEL OF INF-GAMMA FROM MEMORY  
EFFECTOR CELLS DURING THE FIRST PHASE OF A PROTECTIVE IMMUNE  
RESPONSE.  
CC -!- SUBCELLULAR LOCATION: Secreted  
CC -!- SIMILARITY: BELONGS TO THE ESAT6 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: U34848; AAC44033.1; -  
DR EMBL: X79562; CAAS56099.1; -  
DR EMBL: AL022120; CAAL1967.1; -  
DR EMBL: AF007190; AAK48357.1; -  
DR EMBL: AF004671; AAC83446.1; -  
DR TIGR: MT3989; -  
DR TuberculList; Rv3875; -  
KW Antigen; Complete proteome.  
FT INIT\_MET 0  
SQ SEQUENCE 94 AA; 9773 MW; 19245B0EC478BC84 CRC64;

alignment\_scores:  
Quality: 474.00 Length: 94  
Ratio: 5.043 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-462-480-1 x ESA6\_MYCTU ..

Align seg 1/1 to: ESA6\_MYCTU from: 1 to: 94

```

863 ACAGGACGACGAGTGAATTTTCGGGTATCGAGCGCGCGCAAGCAAT 912
1 ThrGluGlnGlnTTPAsnPheAlaGlyIleGluAlaAlaSerAlaI 17
913 CCAGGGAAATGTCAGTCCATTCCCTTCCTTCAGAGGGAGAGCAGT 962
17 eGlnGlyAsnValThrSerIleHisSerLeuLeuAspGluGlyLysGlns 34
963 CCCTGACCAAGCTCGACGCGCGGTGGGGGTAGCGGTTCGGAGCGGTAC 1012
34 erLeuThrLysLeuAlaAlaIlePheGlyGlySerGlySerGluAlaIyr 50
1013 CAGGTGTCCAGCAAAATGGACGCCAGCTACCGAGCTGAACAACGC 1062
51 GlnGlyValGlnGlnLysTyrAspAlaThrAlaThrGluLeuAsnAsnAl 67
1063 GCTCAGAACCTGCGCGGGAGGATCAGGAGCGCGGTTCAGCAATGGCTT 1112
67 aLeuGlnAsnLeuAlaAlaThrIleSerGluAlaGlyGlnAlaMetAlas 84
1113 CGACGGAAGCAAGCTCACTGGGATGTTGCA 1144
84 erThrGluGlyAsnValThrGlyMetPheAla 94

```

seq\_name: SwissProt\_40:YHL1\_EBV

seq\_documentation\_block:

```

ID YHL1_EBV STANDARD; PRT; 660 AA.
AC P03181;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Hypothetical BHFL1 protein.
OS Epstein-barr virus (strain B95-8) (Human herpesvirus 4).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OX NCBI_TaxID=10377;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84270667; PubMed=6087149;
RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
RA Tuffnell P.S., Barrell B.G.;
RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RL Nature 310:207-211(1984).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; V01555; ; NOT_ANNOTATED_CDS.
DR PIR; A03742; Q0BES3.
KW Hypothetical protein; Early protein; Repeat.
FT DOMAIN 149 648 4 x 125 AA TANDEM REPEATS.
FT REPEAT 149 273 1.
FT REPEAT 274 398 2.
FT REPEAT 399 523 3.
FT REPEAT 524 648 4.
SQ SEQUENCE 660 AA; 66244 MW; 86DA1D67A37152A2 CRC64;

```

alignment\_scores:  
Quality: 212.50 Length: 560  
Ratio: 0.944 Gaps: 29  
Percent Similarity: 40.179 Percent Identity: 25.000

alignment\_block:  
US-09-462-480-1/rev x YHL1\_EBV

```

Align seg 1/1 to: YHL1_EBV from: 1 to: 660
1202 AGGTGCAACTCGCCGATCCCGTGTTT.....CGCTATTCTACGCGAAC 1159
||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
136 ArgGlyProArgProHisProAlaPheGlnValGlnTrpSerAlaArgAs 152
TCGGCGTTGCCCTATGCGAACATCCAGTGACGTTGCCCTTCGGTCGAAGC 1109
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
152 nProGlyCysPro...ArgThrTrpArgArgArgSerGlyAlaGlnArg 168
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1108 CATTCCTGACCGCGGTTCGCTGCGCGCCAGGTTCTGCACCGCGGT 1059
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
168 lyHisPro.....ProProGlyAlaGlyGlnArg 177
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1058 TGTTTCAGCTCGGTAGCCGTCGCTGCCCATTTTGTCTGCACACCCCTGGTAC 1009
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
178 ProSerGlyProThrGlyArgProAlaAlaProGlyAlaProGlyTh 194
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1008 GCCT.....CCGAACCCGCTACC 992
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
194 rProAlaAlaProGlyProGlyGlyAlaAlaValProSerGlyAlaI 211
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
991 GCCCAGGCGCGTGGGACGTTGGTCAGGACTGCTTCCCTCCCTCGTCAAGA 942
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
211 hrProHisProGluArgGly...SerGlyProAlaAspProProAlaAla 226
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
941 GGAATGAATGACGTGACATTTCCCTGCTGCGCTTCCG..... 900
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
227 Ala.....ArgLeuProProGluAr 233
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
899 ....CGGCCTCGATACCCGCGAAATTCACCTGCTCTGTCATGTTTTT 854
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
233 gGlnGluProArgLeuProGlnAspLeuAlaAlaGlnArgCysProA 250
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
853 GCTCGGTTCTTTTCGTATTAGCGGTGAGAGCCCATTT..CCGAGGACA 805
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
250 laGlyProProThrArgSerGlyAlaAlaGlnArgThrHisArg 266
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
804 CGCCTGCTGCTGCTCCTCGTCGG..... 781
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
267 ArgProProGlyCysProArgSerAlaArgAsnProGlyCysProArgTh 283
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
780 .....CCCTCGAGTATTG 768
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
283 rTrpArgArgSerGlyAlaGlnArgGlyHisProProGlyAlaG 300
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
767 GACCGCGCGCTGACGAATATTCGTCGAGATCTCGTCGAGTTCCTGCTTCT 718
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
300 lyGlnArgPro.....SerGlyProThrGlyArgProAlaAla 313
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
717 GCTTATTGGCTGCTTCTTTGGAAGGACACACCGCGCTGGCGCGCGCTC 668
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
314 Pro.....GlyAlaProGlyThrProAlaAlaProGlu 324
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
667 CCG.....CCGCGCC 657
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
324 yProGlyGlyAlaAlaValProSerGlyAlaThrProHisProGluA 341
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
656 GCGCCACTGGCCCT.....GCAACGAACCTGCGC..... 628
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
341 rGlySerGlyProAlaAspProProAlaAlaAlaArgLeuProProGlu 357
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
628 ..... 628
358 ArgGlnGluProArgLeuProGlnAspLeuAlaAlaGlnArgCysPr 374
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
627 ....TCGACTCCACCTGCTGATCTGGTGTTCAGTCCCGG..... 589
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
374 oAlaGlyProProThrArgSerGlyAlaAlaGlnArgThrHisA 391
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
588 .....AGATCCGCTCGAAT..... 574
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
391 rArgProProGlyCysProArgSerAlaArgAsnProGlyCysProArg 407
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```





```

US-09-462-480-1/rev x WAIP_HUMAN
Align seg 1/1 to: WAIP_HUMAN from: 1 to: 503

1268 GCCCCATGACAACTCTCAGAGTCCGCTCAACGATATAACACAGAGAAA 1219
|||||
106 AlaGlyMetProLysLeuArgSerThrAlaAsnArgAspAsnAspSerG1 122
|||||
1218 GGGGAG.....ACGACGGAAGTTCGAACCTCGCCCG 1187
|||||
122 yGlySerArgProProLeuLeuProProGlyGlyArgSerThrSerAlaL 139
|||||
1186 ATCCGGTGTTCGTTATCTACGGAACCTCGCGTTCGCCATCGCAACA 1137
|||||
139 ysPro...PheSerProSerGlyProGlyArgPheProValProSer 154
|||||
1136 TCCAGTACGCTGCTTCGTCGAGCAATGCTGACCGCTTCGCTG 1087
|||||
155 ProGlyHisArg.....SerGlyProProGluProGlnArgAsnArg.. 168
|||||
1086 ATCGTCCCGCGCAGGTTCTGACGCGGTTGTTACGTCGGTAGCGTGGC 1037
|||||
169 .....MetPro.....ProProA 173
|||||
1036 GTCCCATTTTCTCGTGGACACCTGTGTAG.....CTCCGAACGCTAC 993
|||||
173 rgProAspValGlySerLysProAspSerIleProProValProSer 189
|||||
992 CGCCCGAGCGCTCGCA...GCTTGGTCAGGACTGCTTCCCTCGTCA 946
|||||
190 ThrProArgProIleGlnSerSerLeuHisAsnArgGlySerProProVa 206
|||||
945 AGGAGGAATGAATGGAGCTGACATTTCCCTGATTGCTGCGCGG. 897
|||||
206 lProGly.....GlyProArgGlnProSerP 215
|||||
896 .....CTCGATACCGCGAATTCACATGCTGCTGTGCA 861
|||||
215 roGlyProThrProProProProGlyAsnArgGlyThrAlaLeuGly 231
|||||
860 TGTTTTTCCTCGTTCTTTTCGATTAGCGGTGAGAGCCATTCG 811
|||||
232 GlyGlySerIleArgGlnSerProLeuSerSerSer..SerProPheSer 247
|||||
810 AGGACGAGCGCTGCT.....GCTGC 791
|||||
248 AsnArgProProLeuProProThrProSerArgAlaLeuAspAspLysPr 264
|||||
790 TCCTCGTCCGCTCGAGTATTGGACGCGGCTGACGAATATTCGTGA 741
|||||
264 oProProProProProProValGlyAsnArgPro.....SerIleH 278
|||||
740 GATCTCGTCGAGTCTCTGCTTCTGCTTATGGCTGCTTCTTGGAGCGCA 691
|||||
278 IsArgGluAlaValPro.....ProPro 285
|||||
690 CCACCG.....CGGCTGGCGCGCTCCCGCGCGCGCGCCAC 650
|||||
286 ProProGlnAsnAsnLysProProValProSerThrProArgProSerAl 302
|||||
649 TGGCCCTCAAGAACCTCGCGTACCTCCACCTGCTGATCTGGGTTT 600
|||||
302 aProHisArgProHisLeuArgProProProProSerArgProGly.... 317
|||||
599 CAGGTGCGCGGAGATCCGCTCGAATTACCTCCCTCCCTGCGGAGGTAG 550
|||||
318 .....ProProPro..... 320
|||||
549 CGGCATCGGTCTTCATCTCTGCGATGCTGGACTACTTCTCTTTACCT 500
|||||
321 ..... 321
|||||
499 TCCTCGCCAAAATGTTGCAAGTCTTCCGG.CCCGGGTGGCGGGAAGTC 451

```

```

|||||
321 uProProSerSerGlyAsnAspGluThrProArgLeuProGlnArgA 338
|||||
450 TGTTCATTACGGAGCTCACCAGTCTCTCTTCGTCCTCCAGTCGCT 401
|||||
338 snLeuSerLeuSer.....SerSerThrProProLeuProSer..... 350
|||||
400 CGTCTCTTCTTCACGCTCTCGCGAGCGGTGCGCGCGACACGACCC 351
|||||
351 .....ProGlyArgSerGlyProLeuProProProPr 361
|||||
350 GGGCTGGTGGAGCCCGGATTGGACACCTGGCCCATCGCTCCCGGACC 301
|||||
361 oSer.....GluArgProProProValArgAspP 372
|||||
300 CACCGAGCGCGCCACCGCTCACCGACGATCCGCGCAACGCGCGCA 251
|||||
372 roProGlyArgSerGlyProLeuPro.....ProPro... 382
|||||
250 TCACGAGGGGCAACCGCTTTTCGATCAGCTGAGACATCAGCGCGGTG 201
|||||
383 .....ProProValSerArgAsnGlySerThrSerArgAlaLe 395
|||||
200 CGGTCAACGACCCACCTGCGCCA.....GGTAGGACTCCGCGCG 160
|||||
395 uProAlaThr..ProGlnLeuProSerArgSerGlyValAspSerProAr 411
|||||
159 CAGCAGCGCGCGCGCTGGGCGCTGATCCACCGACCGCGGATGTT 110
|||||
411 gSerGlyProArgProProLeuProProAspArgPro..... 423
|||||
109 TCGACAGCGGACTGCTGTCGCGAGCAGCGCCATCTGCGCGGCTTCTCTCGTCG 60
|||||
424 .....Ser 424
|||||
59 GCTGGTTCGCGCGCGCGCTGCGCGCCACCC 30
|||||
425 AlaGlyAlaProProProProSer 434
|||||
seq_name: SwissProt_40:ESA6_MYCLE

seq_documentation_block:
ID ESA6_MYCLE STANDARD; PRT; 95 AA.
AC Q50206; O33083;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 6 kDa early secretory antigenic target homolog (ESAT-6-like protein)
DE (L-ESAT).
GN ESAT6 OR ESX OR L45 OR ML0049 OR MLCB628.12C.
OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RX Wieles B., Notenboom T., Naafs B., Offringa R., Ottenhoff T.;
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
[2]
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltham T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).

```

```
CC -1- SUBCELLULAR LOCATION: Secreted (By similarity).
CC -1- SIMILARITY: BELONGS TO THE ESA6 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X90946; CAA62441.1; -
DR EMBL; Y14967; CAA75200.1; -
DR EMBL; AL583917; CAC29557.1; -
DR Leproma; ML0049; -
KW Complete proteome.
FT CONFLICT 55 55 R -> Q (IN REF. 1).
FT CONFLICT 90 90 M -> T (IN REF. 1).
SQ SEQUENCE 95 AA; 10465 MW; B1526F78CB2AB8A1 CRC64;

alignment_scores:
  Quality: 186.00      Length: 91
  Ratio: 2.447        Gaps: 0
  Percent Similarity: 83.516      Percent Identity: 36.264

alignment_block:
US-09-462-480-1 x ESA6_MYCLE

Align seg 1/1 to: ESA6_MYCLE from: 1 to: 95

869 CAGCAGTGGATTTCGGGGGTATCGAGCGCGGCAAGCGCAATCCAGG 918
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3 GlnAlaTrpHisPheProAlaLeuGlnGlyAlaValAsnGluLeuGlnG 19

919 AAATGTCAGTCCATTCATCCCTCTTGTGAGGAGGGAAGCATGCCGTA 968
|::: :: || || ||||| ||||| ||||| ||||| ||||| |||||
19 ySerGlnSerArgIleAspAlaLeuLeuGlnGlnCysGlnGluSerLeu 36

969 CCAAGCTCGCAGCGCGTGGGGCGGTAGCGGTTCGGAGGCGTACCGGGT 1018
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
36 hrLysLeuGlnSerSerTrpHisGlySerGlyAsnGluSerTyrSer 52

1019 GTCCAGCAAAATGGGACGCCAGCGGTACCGAGCTGAACACGCGCTGCA 1068
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
53 ValGlnArgArgPheAsnGlnAsnThrGluGlyIleAsnHisAlaLeu 69

1069 GAACCTGGCGGAGCATCAGCGAGCGGTGTCAGCAATGGCTTCGACCG 1118
::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
69 yAspLeuValGlnAlaIleAsnHisSerAlaGluThrMetGlnGlnThr 86

1119 AAGGCAACGTCACCTGGGATGTC 1141
|:::| ||||| ||||| ||||| ||||| ||||| |||||
86 luAlaGlyValMetSerMetPhe 93

seq_name: SwissProt_40:CF10_MYCLE

seq_documentation_block:
ID CF10_MYCLE STANDARD; PRT; 99 AA.
AC Q33084;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 16 kDa culture filtrate antigen cfp10 homolog.
GN MLO050 OR MLCB628.13C.
OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
ON NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RA Eiglmeyer K., Garnier T., De Rossi E., Fsihi H., Cole S.T.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DDBJ databases.
```

```
RN
RP SEQUENCE FROM N.A.
RX STRAIN=TN;
RA MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eiglmeyer K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajadream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).
CC -1- SIMILARITY: BELONGS TO THE ESA6 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y14967; CAA75210.1; -
DR EMBL; AL583917; CAC29558.1; -
DR Leproma; ML0050; -
KW Complete proteome.
FT INIT_MET 0 0 BY SIMILARITY.
SQ SEQUENCE 99 AA; 10833 MW; 859B484F7EFE5A8A CRC64;

alignment_scores:
  Quality: 182.00      Length: 99
  Ratio: 2.333        Gaps: 0
  Percent Similarity: 78.788      Percent Identity: 39.394

alignment_block:
US-09-462-480-1 x CF10_MYCLE

Align seg 1/1 to: CF10_MYCLE from: 1 to: 99

528 CGACGAGTGAAGACCGATGCGCTACCTCGGCGGAGGAGGAGGTAATTT 577
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1 AlaGluMetIleThrGluAlaAlaIleLeuThrGlnGlnAlaAlaGlnPh 17

578 CGACCGGATCTCCGGCGACCTGAACCCAGATCGACGAGTGGAGTCGA 627
|:::| ||||| ||||| ||||| ||||| ||||| ||||| |||||
17 eAspGlnIleAlaSerGlyLeuSerGlnGluArgAsnPheValAspSerI 34

628 CGGCGAGTTCGTTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCC 677
:::|::| ||||| ||||| ||||| ||||| ||||| ||||| |||||
34 leGlyGlnSerPheGlnAsnThrTrpGluGlyGlnAlaAlaSerAlaAla 50

678 CAGGCGCGGTGGTGGCGCTTCCAAAGACGACGCCAATAAGCAGAAGCAGA 727
:::|::| ||||| ||||| ||||| ||||| ||||| ||||| |||||
51 LeuGlyAlaLeuGlyArgPheAspGluAlaMetGlnAspGlnIleArgG 67

728 ACTCGAGGAGATCTCGAGCAATATTCGTACGGCGGCGGTCCATACTCGA 777
:||||:| ||||| ||||| ||||| ||||| ||||| ||||| |||||
67 nLeuGluSerIleValAspLysLeuAsnArgSerGlyGlyAsnTyrThrL 84

778 GGGCGGACGAGGAGCAGCAGCGCGCTGCTCGCAAAATGGGCGTTC 824
:::|::| ||||| ||||| ||||| ||||| ||||| ||||| |||||
84 yThrAspAspGluAlaAsnGlnLeuLeuSerSerLysMetAsnPhe 99

seq_name: SwissProt_40:VNUA_PVKA

seq_documentation_block:
ID VNUA_PVKA STANDARD; PRT; 1733 AA.
AC P33485;
DT 01-FEB-1994 (Rel. 28, Created)
```

DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 01-FEB-1994 (Rel. 28, Last annotation update)  
 DE Probable nuclear antigen.  
 OS Pseudorabies virus (strain Kaplan) (PRV).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Alphaherpesvirinae; Varicellovirus.  
 OX NCBI\_TaxId=33703;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91021039; PubMed=2171211;  
 RA Vlcek C., Kozmik Z., Paces V., Schirm S., Schwyzler M.;  
 RT "pseudorabies virus immediate-early gene overlaps with an oppositely  
 oriented open reading frame: characterization of their promoter and  
 enhancer regions.";  
 RL Virology 179:365-377(1990).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; M34651; AAA47471.f;  
 DR PIR; B45344; B45344.  
 FT DOMAIN 112 117 POLY-THR.  
 FT DOMAIN 179 1733 GLY-RICH.  
 FT DOMAIN 192 196 POLY-SER.  
 FT DOMAIN 271 298 POLY-PRO.  
 FT DOMAIN 304 308 POLY-ARG.  
 FT DOMAIN 883 889 POLY-GLY.  
 FT DOMAIN 1398 1405 POLY-GLY.  
 FT SEQUENCE 1733 AA; 172166 MW; 0C8CD8B8475BB5E2 CRC64;  
 SQ

alignment\_scores:  
 Quality: 172.50 Length: 504  
 Ratio: 0.757 Gaps: 29  
 Percent Similarity: 45.238 Percent Identity: 27.183

alignment\_block:  
 US-09-462-480-1 x VNUA\_PRVKA ..  
 Align seg 1/1 to: VNUA\_PRVKA from: 1 to: 1733

```

31 GTGGCGCGCCGCGC.....GGCGG 50
||||||| |||
840 ValGlyGlyGluGlyArgLeuGlyGlyProArgValGlyLeuAlaG 856

51 CAACCCAGCCGACGAGCGCGCAGATGGCGCTCTCGGCACGATC 100
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
856 yArgSPAlaAlaAlaAlaAlaValGlyArgGlyValLeuGlyHisGlyP 873

101 CG...CTGTGAACATCCGCTG.....GCTGTGGATCAGGC 135
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
873 roGluArgAlaProGluProValLeuGlyGlyGlyGlyGlyGly 889

136 CCCAGCGCGCGCGCGCTG..... 156
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
890 GlnGlnArgGlySerGlyValArgSerGlyProGluSerGluGlyAlaAl 906

157 .....C 157

906 aLeuAlaProGlyProProValLeuPheValAlaValAlaVal 923

158 TCGCGCGCGAG.....TCGCTACTT 177
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
923 alProAlaGluGlyArgAlaGlyGluProLeuValLeuAlaValPro 939

178 GCGCAGGTGGGTGGTTGACCCGCGCGCTGTGTGTGTGTGTGTGTG 227
||||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
940 GlyAlaAlaGly.....ProGlyArgAlaAlaLeuLeuLeu 951

```

```

228 AAGCCGGTTCCTCCCTCGGTGATGCGCGCGCTGTTCGCCGATCGTCGG 277
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
951 uAlaProLeuGlyArgTrpValArg...AlaGlyGlyGlyGlyAlaGlyV 967

278 TGACGGGTGGCGCGCTCCCGGTGGTCCGGGAGCG..... 312
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
967 aLaAlaGlyGlyAlaGlyGluAlaGlyLeuGlyAlaGlyAlaGlyLeuGly 983

313 ATGGCCAGAGGTTCGCAATCGCGCGCTCCAC.CAGCCCGGTGTGTGTCG 361
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
984 AlaGlyAlaGlyLeuGlyAlaGlyGlyAlaGlyGlyProGlyAlaGlyG 1000

362 CGCCGCGCACCTCGCGCAGGCGGTGAAGAAGACGA..... 398
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1000 uAlaGly.....GlyGlyAlaAlaArgArgArgArgArgTrpA 1014

399 .....CGAGCAGACTGGCAGGAGA 419
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1014 sPAspGluAlaGlyLeuLeuGlyProGluArgGlyGlnAlaGlyArgGly 1030

420 GGACGACTGGTGAGCTCCCGTAATGACACAGACTTCCC.....GGCCA 463
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1031 LeuArgGlyProGlyProArgGlyGlyLeuGlyGluProGlyArgGlyH 1047

464 CCCGGCGCGGAAGACTTCCCAACATTTTGGCGAGGAGGTAAA..... 506
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1047 sValGlyArg.....GlyGluGlyArgGlyValG 1058

507 .....GAG 509
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1058 LyProGlyGlyLeuAlaGlyAlaGlyProValHisAlaValAlaHisGln 1074

510 AGAAGTAGTCCAGATGGCAGATGAAGACCGATGCCCTACCTCCGCG 559
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1075 ArgArgHisGlyAla...GlyAspGluGlyAspArgValArgGlyLeuPro 1090

560 GCAGGAGCGAGTAATTTTCGAGCGGATCTCCGGCGACCTGAAACCCAGA 609
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1091 ProLeuGlyArgAlaGlyProGlyAsp...ArgValAlaGluArgGluG 1106

610 TCGACGAGTGGAGTCGACGCGCAGTTCGTT.....GCAGGCGCAG 650
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1106 nArg.....GlyArgHisLeuLeuGluAlaGlyGlyProG 1118

651 TGGCGCGCGC.....GGCGGCGCGCGCGCGCGCGCGTGT... 691
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1118 LuGlyGlyArgGlyAlaGlyGlyArgGlyGlnProGluArgAlaGlyGln 1134

692 ...GCGCTTCCAAGACGACCAATAAGCAGAAGCAGGAACCTCGACGAGA 738
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1135 GlnAlaLeuGluAspAlaAlaAlaGlyGlnAspAlaGlyValArgGln.. 1150

739 TCTCGAGCAATATTCTCGTCCGCGCGCTCCAATCTCGAGGCGCGCAGAG 788
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1151 .LeuAlaGlyHisAlaAlaGlyLeuArg.....GlyGlyGluGlyG 1164

789 GAGCAGCAGCAGCGCGTCTCTCGCAATGGCTTCTGACCCGCTAATAC 838
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1164 LyAlaAspAlaGlyAlaGlyGlyLeuAspGlyArgLeuProGly..... 1178

839 GAAAGAAACGAGCAAAACATCAGCAGCAGCAGTGGAAATTCGCGGG 888
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1179 .....AlaG 1180

889 TATCAGCGCGCGCAAGCGCAATCCAGGAAATGTACATTCATT 938
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1180 yValArgGlyAlaAlaArgValGlyHisValGlyValGlyProAlaGluA 1197

939 CCTCTCTTGACGAGGGAAGCAGTCCCTGACCAAGCTCGCAGCGCGCTGG 988
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1197 laLeuGlnAspGluGly.....LeuLeuGlyAlaIleValAlaAla 1211

```



[illegible]

```

677 hrPhElysProGlySerProThrValGlyProGlyProLeuProProAlaA 693
86 AGGCCCATCTCGGGGCTTCTCTCGGTGGTGCCGCCGCCTGGCC 37
   |||      :|||  |||::  |||||||:::
694 GlyPro.....SerSerLeuSerSerLeuProProProAlaA 707
36 GCCCACC 30
   |||||
707 aProthr 709

seq_name: SwissProt_40:IRS2_HUMAN
seq_documentation_block:
ID IRS2_HUMAN STANDARD; PRT; 1324 AA.
AC Q9Y4H2;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Insulin receptor substrate-2 (IRS-2).
GN IRS2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97460123; PubMed=9312143;
RA Ogiwara T., Isobe T., Ichimura T., Taoka M., Funaki M., Sakoda H.,
RA Onishi Y., Inukai K., Anai M., Fukushima Y., Kikuchi M., Yazaki Y.,
RA Oka Y., Asano T.;
RT "I4-3-3 protein binds to insulin receptor substrate-1, one of the
RT binding sites of which is in the phosphotyrosine binding domain." ;
RL J. Biol. Chem. 272:25267-25274(1997).
CC -!- FUNCTION: MAY MEDIATE THE CONTROL OF VARIOUS CELLULAR PROCESSES
CC BY INSULIN.
CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 PTB DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collabora-
CC between the Swiss Institute of Bioinformatics and the EMBL outst-
CC the European Bioinformatics Institute. There are no restrictions
CC use by non-profit institutions as long as its content is in
CC modified and this statement is not removed. Usage by and for com-
CC entities requires a license agreement (See http://www.isb-sib.ch/an-
CC or send an email to license@isb-sib.ch).
CC -----
EMBL; AB000732; BAA24500.1; -.
DR HSPSP; P35568; IIRS.
DR MM; 600797; -.
DR InterPro; IPR002404; Insulin_Recep_S-1.
DR InterPro; IPR001849; PH.
DR Pfam; PF02174; IRS; 1.
DR Pfam; PF00169; PH; 1.
DR PRINTS; PR00628; INSULINSR1.
DR SMART; SM00233; PH; 1.
DR SMART; SM00310; PTBI; 1.
DR PROSITE; PS50003; PH_DOMAIN; 1.
KW Phosphorylation.
FT DOMAIN 16 144
FT FT 190 303
FT MOD_RES 540 540
FT FT 653 653
FT MOD_RES 675 675
FT FT 919 919
FT MOD_RES 978 978
FT FT 1253 1253
FT MOD_RES 19 28
FT DOMAIN

```

```

642  GCAACGAACCTGCGCTCGACTCCACCTGGTGATCTGGGTTTCAAGTCG  593
      ::::::::::: |||||
1026  SerSerLeuGlnProProPro.....  1034

592  CGGAGATCCGCTCGAATTAACCTGCCTCTCGCCGAGGTAGCGCATC  543
      ::::::::::: |||
1035  .....ProProAlaProGlyGluLeuTyrA  1043

542  GGTCTTCATCTCTGCCATGCTGGACTACTTTCTCTCTTTACCTTCCTCGC  493
      || ::||| |||
1043  rr.....LeuPro.....ProAla  1047

```

1043 rg.....LeuPro.....Prola 104  
492 CAAATGTTGCAAGTCTTCCGGCCCGGTGCCCGGAAGTCTGTGTCA 443  
1048 SerAlaValAlaThrAlaGlnGlyProGlyAlaAla..... 1059

442 TTACGGAGCTCACCAGTCGTCCTCTTTCGTCCCGTCGTCGTCGTCCT 393  
||||| ||||| : : : : : : : : : : : :  
1060 .....SerSerLeuSerSerAspThrGlyAspAsnGly 1070

1070 ..... SerSerLeuSerSerAspIrrnGlyAspSnGly 1070  
392 TCCTTCAGCCTCCTCGCGAGCGACGGTCCGCCGCACCAGACCCGGCGTGTT 343  
::: :::: ||| ||||| |  
1071 AspYrThrGlutMetAlaPheGlyValAlaAlaThr..... 1082

392 TCTTCACGCTCCTCGGAGGGTCCGGGCGGCACCAAGCCCCGGGTGGT 343  
 ... ||| ||||| |||||  
 1071 AspTyrThrGluMetAlaPheGlyValAlaAlaThr ..... 1082  
 342 GGAGCGCGGATGCTGGAAACCTGGCCCATGTCTCCCGGACCACCGGAG 293  
 ... ||| |||  
 1083 .....ProProGlnProIle.....A 1088

292 GGGGCCACCGCGTCACGACATCATCGGCACAGCGCGGCATCATCCGAG 243  
 |||||  
 1088 IaAlaProPro...LysProGluAlaAlaArgValAlaSerProThrSer 1103  
 |||||  
 242 GGGGCCACCGCGCTTTTCATCATCGTCAGACATCATCAGCGCGCTG ..... 201  
 |||||  
 1104 GlyValLysArgLeuSerLeuMetGluGlnValSerGlyValGluAlaPh 1120

200 ... CGGCTCAACGACCACCTGCGCCAGTAGCTAGCGCTCCGCGCGCAGCA 155  
 1120 eLeuGlnAlaSerGlnProAspPro.....HisA 1131  
 154 GGCCGCGCGCGCTGGGGCCTCATCCACGACCGCGATGTTGCGAC 105  
 1131 rgGlyAlaLysvalleargAlaAspGlnGlyGly..... 1143

104 ACGGACTGTGTCGCGACGACGCCCATCTGCGCGCTCTCCATCCGCTGG 55  
1144 .....ArgArgArgHisSerSerGluThrPheSerThrTh 1156  
54 GTTGCGCGCGCGGTGCGCGCCACCTGGCTGCACAC 18  
1156 rThValThrProValSerProSerPheAlaHisAsn 1168

```

name: SwissProt_40:T2D3_HUMAN
documentation_block:
  T2D3_HUMAN STANDARD; PRT: 1083 AA.
  Q00268; Q99721; Q9BX42; Q9BR40;
  15-JUL-1998 (Rel. 36, Created)
  15-JUL-1998 (Rel. 36, Last sequence update)

```

01-MAR-2002 (Rel. 41, Last annotation update)  
Transcription initiation factor TFIID 135 kDa subunit  
(TAFII135) (TAFII-130) (TAFII130).  
TAF4A OR TAF2C1 OR TAF2C OR TAFII135 OR TAFII130.  
Homo sapiens (Human).  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Theria; Primates; Catarrhini; Homiidae; Homo.  
NCBI\_TaxID=9606;

[1]  
SEQUENCE FROM N.A.  
MEDLINE-97336072; pubMed-9192867;  
Mengus G., May M., Carre L., Chambon P., Davidson I.:  
"Human TAF(II)135 potentiates transcriptional activation by

seq documentation block:

seq\_documentation\_block:

AD	T2D3_HUMAN	STANDARD;	PRT;	1083	AA.
AC	C00268;	Q99721;	Q9BX42;	Q9BR40;	
DT	15-JUL-1998	(Rel. 36,	Created)		
DT	15-JUL-1998	(Rel. 36,	Last sequence update)		
DT	01-MAR-2002	(Rel. 41,	Last annotation update)		
DE	Transcription initiation factor TFIID 135 kDa subunit (TAFII-135)				
DE	(TAFII135) (TAFII-136) (TAFII130).				
GN	TAF4A OR TAF2C1 OR TAF2C OR TAFII135 OR TAFII130.				
OS	Homo sapiens (Human).				
OC	Eukaryota;	Metazoa;	Chordata;	Craniata;	Vertebrata; Euteleostomi;
OC	Mammalia;	Eutheria;	Primates;	Catarrhini;	Hominidae; Homo.

RP SEQUENCE FROM N.A.

RX MEDLINE=97336072; PubMed=9192867;  
RA Mengus G., May M., Carre L., Chambon P., Davidson I.;  
RT "Human TAF(II)135 potentiates transcriptional activation



314 A.....TCGCTCCGGACCCAGCGGAGCGGCCACCGCTCAC 277  
| : : : | | | | | : : : |  
261 caAlaAlaProAlaAlaProProProPro.....ProAlaP 275  
276 CGACATCCGCAACAGCGCGGCATCACCGAGGGGCAACCGCTTT 227  
| : : : | | | | | : : : | | | | : : : |  
275 foAlaThLeuAlaArgProGlyHisProAlaGlyProProhAla 291  
| : : : | | | | | : : : | | | | : : : |  
226 CGATCAGCT.....GAGCATCAGCGCGGTGCGGTCACGACCCA 186  
| | : : : | | | : : : |  
292 AlaProAlaValProProProAlaAlaAlaGlnAsnGlyGlySerAlaG1 308  
185 CTGCGCCAGTAGGACTCCGCGCGAGCAGCGCCCGCGCGCTGGG 136  
| : : : | | | | | : : : | | | | : : : |  
308 yAlaAlaProAlaProAlaProAlaAlaGlyGlyProAla.....G 322  
135 GCCTGATCCACGACGCGGATGTCGACGCGGACTGTCGCGGACCA 86  
| | : : : | | | : : : | | | : : : |  
322 lyValSerGlyGlnProGlyProGlyAlaAlaAlaAlaAlaProAlaPro 338  
85 GGCCCATCTGCGCGCTTCCTCGTGGTGGTTCGCCCGCGCTGCGG 36  
| | : : : | | | : : : | | | : : : |  
339 GlyValysAlaGluSerProlysArgVal.....ValGlnAlaAl 352  
35 CCCACCTGGGTGACAAAG.....ACGTACCTGCTGCA 2  
| | | | : : : | | | : : : | | | : : : |  
352 aProProAlaAlaGlnThrLeuAlaAlaSerGlyProAlaSer 366

seq\_name: SwissProt\_40:DRPL\_HUMAN

seq\_documentation\_block:

ID DRPL\_HUMAN STANDARD; PRT; 1185 AA.  
AC P54259;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Acrophin-1 (Dentatorubral-pallidolysian atrophy protein).  
GN DRPLA.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
[1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Cerebellum, and Brain;  
RX MEDLINE=95144175; PubMed=7842016;  
RA Nagafuchi S., Yanagisawa H., Ohsaki E., Shirayama T., Tadokoro K.,  
RA Inoue T., Yamada M.;  
RT "Structure and expression of the gene responsible for the triplet  
RT repeat disorder, dentatorubral and pallidolysian atrophy (DRPLA).";  
RL Nat. Genet. 8:177-182(1994).  
[2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96262314; PubMed=8965642;  
RA Margolis R.L., Li S.-H., Young W.S., Wagster M.V., Stine O.C.,  
RA Kidwai A.S., Ashworth R.G., Ross C.A.;  
RT "DRPLA gene (atrophin-1) sequence and mRNA expression in human  
RT brain.";  
RL Brain Res. Mol. Brain Res. 36:219-226(1996).  
[3]  
RP SEQUENCE OF 470-725 FROM N.A.  
RC TISSUE=Brain cortex;  
RX MEDLINE=93315145; PubMed=8325628;  
RA Li S.-H., McInnis M.G., Margolis R.L., Antonarakis S.E., Ross C.A.;  
RT "Novel triplet repeat containing genes in human brain: cloning,  
RT expression, and length polymorphisms.";  
RL Genomics 16:572-579(1993).  
CC -!- TISSUE SPECIFICITY: THE LEVELS ARE RELATIVELY HIGH IN THE BRAIN,  
CC OVARY, TESTIS AND PROSTATE. LOWER LEVELS ARE DETECTED IN THE  
CC LIVER, THYMUS AND LEUCOCYTES.  
CC -!- POLYMORPHISM: THE POLY-GLN REGION OF DRPLA IS HIGHLY POLYMORPHIC  
CC (7 TO 23 REPEATS) IN THE NORMAL POPULATION AND IS EXPANDED TO  
CC ABOUT 49-75 REPEATS IN DRPLA PATIENTS. LONGER EXPANSIONS RESULT IN

CC CC EARLIER ONSET AND MORE SEVERE CLINICAL MANIFESTATIONS OF THE  
CC DISEASE.  
CC -!- DISEASE: DEFECTS IN DRPLA ARE THE CAUSE OF DENTATORUBRAL-  
CC PALLIDOLYSIAN ATROPHY, AN AUTOSOMAL DOMINANT NEURODEGENERATIVE  
CC DISORDER CHARACTERIZED BY A LOSS OF NEURONS IN THE DENTATE  
CC NUCLEUS, RUBRUM, GLOBUS PALLIDUS AND LUY'S BODY. CLINICAL FEATURES  
CC ARE MYOCLONUS EPILEPSY, DEMENTIA, AND CEREBELLAR ATAXIA. ONSET OF  
CC THE DISEASE OCCURS USUALLY IN THE SECOND DECADE OF LIFE AND DEATH  
CC IN THE FOURTH.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; D31840; BAA06626.1; -;  
CC EMBL; U23851; AAB50276.1; -;  
CC EMBL; L10377; -; NOT\_ANNOTATED\_CDS.  
CC HSSP; P00651; 1LRA.  
CC MIM; 125370; -;  
CC  
CC InterPro: IPR002951; Atrophin.  
CC PRINTS; PR01222; ATROPHIN.  
CC KW Triplet repeat expansion; Polymorphism.  
CC FT DOMAIN 73 82 SER/GLU-RICH (MIXED CHARGE).  
FT DOMAIN 302 305 POLY-PRO.  
FT DOMAIN 376 382 POLY-SER.  
FT DOMAIN 386 397 POLY-SER.  
FT DOMAIN 442 447 POLY-PRO.  
FT DOMAIN 479 483 POLY-HIS.  
FT DOMAIN 484 497 POLY-GLN.  
FT DOMAIN 504 507 POLY-PRO.  
FT DOMAIN 564 574 POLY-SER.  
FT DOMAIN 704 707 POLY-PRO.  
FT DOMAIN 802 815 ARG/ALA-RICH (MIXED CHARGE).  
FT DOMAIN 816 827 ARG/GLU-RICH (MIXED CHARGE).  
FT DOMAIN 925 934 ARG/GLU-RICH (MIXED CHARGE).  
FT CONFLICT 94 94 MISSING (IN REF. 2).  
FT CONFLICT 333 333 Y -> H (IN REF. 2).  
FT CONFLICT 339 339 M -> I (IN REF. 2).  
FT CONFLICT 541 541 P -> T (IN REF. 3).  
FT CONFLICT 1028 1028 G -> A (IN REF. 2).  
SQ SEQUENCE 1185 AA; 124785 MW; 56C30626731C005 CRC64;

alignment\_scores:

Quality: 165.00 Length: 536  
Ratio: 0.721 Gaps: 35  
Percent Similarity: 42.724 Percent Identity: 25.746

alignment\_block:

US-09-462-480-1/rev x DRPL\_HUMAN ..

Align seg 1/1 to: DRPL\_HUMAN from: 1 to: 1185

1267 CGCGCATGACAACTCTCAGAGTGGCTCAACGTAATAAACACGAGAAAG 1218  
||||| ||| : : : ||| : : : : :  
255 ProProThrThrProIleSerValSerSer.....GlyAlaSerG1 269  
1217 GCGGAGACCGGACGGAAGTTCGACATCCGCGCGATCCCGTTCGCTATTC 1168  
||| ||||| : : : ||| : : : : :  
269 yAlaProThrThrLysProProThrThrProValGlyGlyGlyAsnLeu. 285  
1167 TACCGCAACTCGCGCTTCCCTATCCGACATCCGACGACGTCGCTTC 1118  
||||| ||| ||| : : : ||| : : : : :  
286 .....Pro.SerAlaProProAlaAsnPheProHisValThrProAs 300  
1117 GGTGCAAGCATTCGCTGACCGGCTTCG.....TGATCG 1083  
: : : ||| ||| ||||| : : : : :  
300 nLeu.ProPro....ProProAlaLeuArgProLeuAsnAlaSerAla 315







```

1083 CGATCAGCGAAGCGGTACGC 1104
      |          : : : : |
762 pGlyThr***GlyGluArgGly 769

seq_name: SwissProt_40:CA21_BOVIN

seq_documentation_block:
ID   CA21_BOVIN      STANDARD;          PRT;   1364 AA.
AC   P21JUL5; O62649;
DC   20-01-1986 (Rel. 01, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DI   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Collagen alpha 2(I) chain precursor.
GN   COL1A2.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Aorta;
RX   MEDLINE=98290219; PubMed=9628255;
RA   Shirai T., Hattori S., Sakaguchi M., Inouye
RA   Irie S., Nagai Y., Hori H.;
RT   "The complete cDNA coding sequence for the b
RT   of type I procollagen.";
RL   Matrix Biol. 17:85-88(1998).
RN   [2]
RP   SEQUENCE OF 80-98.
RC   TISSUE=Skin;
RX   MEDLINE=75036115; PubMed=4609475;
RA   Fietzek P.P., Bretkreutz D., Kuehn K.;
RT   "Amino acid sequence of the amino-terminal r
RT   collagen.";
RL   Biochim. Biophys. Acta 365:305-310(1974).
RN   [3]
RP   SEQUENCE OF 95-415, AND REVISION.
RC   TISSUE=Skin;
RX   MEDLINE=76091874; PubMed=173531;
RA   Fietzek P.P., Rexford F.W.;
RT   "The covalent structure of collagen. The am
RT   alpha2-CB4 from calf-skin collagen.";
RL   Eur. J. Biochem. 59:113-118(1975).
RN   [4]
RP   SEQUENCE OF 416-445.
RC   TISSUE=Skin;
RX   MEDLINE=75008198; PubMed=4412529;
RA   Fietzek P.P., Furthmayr H., Kuehn K.;
RT   "Comparative sequence studies on alpha2-CB2
RT   and pig-skin collagen.";
RL   Eur. J. Biochem. 47:257-261(1974).

```

RA Fietzek P.P., Kuehn K.:  
RT "The covalent structure of collagen: amino acid sequence of the N-  
RT terminal region of alpha2-CB3 from rat skin collagen and alpha2-CB3.5  
RT from calf skin collagen.";  
EL Hoppe-Sevler's Z. Physiol. Chem. 355:647-650(1974).  
CC -1- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN  
CC (FIBRILLAR FORMING COLLAGEN).  
CC -1- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.  
CC -1- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND  
CC BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM  
CC HYDROXYAPATITE.  
CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -



```

DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00660; BAND_41_1; FALSE_NEG.
DR PROSITE; PS00661; BAND_41_2; FALSE_NEG.
DR PROSITE; PS50057; BAND_41_3; 1.
DR PROSITE; PS50096; IQ; 3.
DR PROSITE; PS50002; SH3; FALSE_NEG.
DR Myosin; ATP-binding; Actin-binding; Coiled coil; Repeat; SH3 domain;
KW Calmodulin-binding; Disease mutagen; Deafness.
FT DOMAIN 1 1887 HEAD OR MOTOR DOMAIN.
FT DOMAIN 1888 2029 NECK OR REGULATORY DOMAIN.
FT DOMAIN 2030 3530 TAIL.
FT DOMAIN 1323 1350 COILED COIL (POTENTIAL).
FT DOMAIN 1792 1799 ACTIN-BINDING (POTENTIAL).
FT DOMAIN 1902 1924 IQ 1.
FT DOMAIN 1925 1954 IQ 2.
FT DOMAIN 1955 1976 IQ 3.
FT DOMAIN 2867 2953 SH3.
FT DOMAIN 3206 3443 BAND 4.1-LIKE.
FT NP_BIND 1315 1322 ATP (POTENTIAL).
FT VARIANT 2111 2111 N -> Y (IN DNB3; FAMILY FROM BENGKALA)
FT VARIANT 2113 2113 I -> F (IN DNB3; INDIAN FAMILY).
FT VARIANT 2113 2113 /FTId=VAR_010303.
FT VARIANT 2113 2113 /FTId=VAR_010304.
SQ SEQUENCE 3530 AA; 395171 MW; 3D103923D4BCBE4A CRC64;

alignment_scores:
Quality: 163.50 Length: 481
Ratio: 0.790 Gaps: 27
Percent Similarity: 43.035 Percent Identity: 26.611

alignment_block:
US-09-462-480-1/rev x MY15_HUMAN ..
Align seg 1/1 to: MY15_HUMAN from: 1 to: 3530

1262 ATGACAACCTCTCAGAGTCGCGCTCAACGTATATAACACGAGAAAGGGCGA 1213
      ::::: ||| |||||::: |||
535 LeuThrProArgGlnArgAsnLeuGlnArgAlaLeuSerAlaPheGlyAl 551
      ::::: ||| |||||::: |||
1212 GACCGACGAGAGTGCAGTCCGCCGAT.....CCCGTGTTCGCT 1172
      ||| ::::: |||||::: ||||| |||
551 aHisArgGlyLeuGlyPheGlyProGluPheGlyArgProValProArgP 568
      ::::: ||| |||||::: |||||
1171 ATTCTACCGCAACTCGCGTTCGCCCTATCGCAACATCCAGTGCAGTTGC 1122
      ::::: ||| |||||::: |||||
568 roAlaThrSerLeuAlaArgPheLeuLysLysLeuSer..... 581
      ::::: ||| |||||::: |||||
1121 CTTCGGTCCGAACCATTCGCTGACCGCGTTCGCTGATCGTCC..... 1080
      ::::: |||||::: ||||| |||||
582 ...GluLysLysProIleAla...ArgLeuArgGlySerGlnLysThrAr 596
      ::::: |||||::: ||||| |||||
1079 .GCGCAGGTTCTGACGCGGTGTTCAGTCGCGTAGCGGTGGCGTCCGA 1031
      ||| ||| |||||::: |||||
596 gAlaGlyGlyProAlaValArgGluAlaAlaTyrLysArgPheGlyTyrL 613
      ::::: ||| |||||::: |||||
1030 TTTTTCGTGGA.....CACCCTGGTACCGCTCCGGAACCGCTAC 993
      ::::: ||||| ||||| |||||
613 ysLeuAlaGlyMetAspProGluLysProGlyThrProIleValLeuArg 629
      ::::: ||||| ||||| |||||
992 CGCCCCAGCGCCGCTCGAGCTTGGTCAGGACTGCTTC....CCCTCGTC 947
      ||| ::::: ||||| |||||
630 ArgAlaGlnProArgAlaArgSerSerAsnAspAlaArgArgProAla 646
      ::::: ||||| ||||| |||||
946 AAGGAGGAAGTAATGGACGTGACATTTCCCTGGATTCGGCTGCGCGGG 897
      : :::: ||| |||||::: |||
646 aProGlnProAlaProArgThrLeuSerHisTrpSerAlaLeuLeuSerP 663
      ::::: ||| |||||::: |||||
896 CCTCGATACCCCGCAATTCACACTGCTGCTGTCATGTTTTGCTCGT 847
      ::::: ||| |||||::: |||||
663 roProValProProArg..... 668
      ::::: ||||| ||||| |||||
846 TTCTTTTCGTATTACGGGTCAAGAGCCATTTCGGAGGACAGCCGTCG 797

```



RC TISSUE=Placenta;  
RX MEDLINE-94375010; PubMed=8088784;  
RA Christiano A.M., Hoffman G.G., Chung-Honet L.C., Lee S., Cheng W.,  
RA Uitto J., Greenspan D.S.;  
RT "Structural organization of the human type VII collagen gene (COL7A1),  
RT composed of more exons than any previously characterized gene.";  
RL Genomics 21:169-179(1994).  
RN [8]  
RP REVIEW ON DEB VARIANTS.  
RX MEDLINE-98041696; PubMed=9375848;  
RA Jaervikallio A., Pulkkinen L., Uitto J.;  
RA "Molecular basis of dystrophic epidermolysis bullosa: mutations in  
RT the type VII collagen gene (COL7A1).";  
RL Hum. Mutat. 10:338-347(1997).  
RN [9]  
RP VARIANT RDEB LYS-2798.  
RX MEDLINE-93291877; PubMed=8513326;  
RA Christiano A.M., Greenspan D.S., Hoffman G.G., Zhang X., Tamai Y.,  
RA Lin A.N., Dietz H.C., Hovnanian A., Uitto J.;  
RT "A missense mutation in type VII collagen in two affected siblings  
RT with recessive dystrophic epidermolysis bullosa.";  
RL Nat. Genet. 4:62-66(1993).  
RN [10]  
RP VARIANT DDEB SER-2040.  
RX MEDLINE-94224777; PubMed=8170945;  
RA Christiano A.M., Ryyanen M., Uitto J.;  
RT "Dominant dystrophic epidermolysis bullosa: identification of a  
RT Gly-->Ser substitution in the triple-helical domain of type VII  
RT collagen.";  
RL Proc. Natl. Acad. Sci. U.S.A. 91:3549-3553(1994).  
RN [11]  
RP VARIANT PEB-DDEB CYS-2623.  
RX MEDLINE-96081220; PubMed=8541842;  
RA Christiano A.M., Lee J.Y.-Y., Chen W.J., Laforgia S., Uitto J.;  
RT "Pretibial epidermolysis bullosa: genetic linkage to COL7A1 and  
RT identification of a glycine-to-cysteine substitution in the triple-  
RT helical domain of type VII collagen.";  
RL Hum. Mol. Genet. 4:1579-1583(1995).  
RN [12]  
RP VARIANT DDEB ARG-2043.  
RX MEDLINE-95164985; PubMed=7861014;  
RA Christiano A.M., Moricone A., Paradisi M., Angelo C., Mazzanti C.,  
RA Cavalieri R., Uitto J.;  
RT "A glycine-to-arginine substitution in the triple-helical domain of  
RT type VII collagen in a family with dominant dystrophic epidermolysis  
RT bullosa.";  
RL J. Invest. Dermatol. 104:438-440(1995).  
RN [13]  
RP VARIANTS RDEB AND DDEB.  
RX MEDLINE-96220218; PubMed=8644729;  
RA Christiano A.M., McGrath J.A., Tan K.C., Uitto J.;  
RT "Glycine substitutions in the triple-helical region of type VII  
RT collagen result in a spectrum of dystrophic epidermolysis bullosa  
RT phenotypes and patterns of inheritance.";  
RL Am. J. Hum. Genet. 58:671-681(1996).  
RN [14]  
RP VARIANT RDEB ARG-2575.  
RX MEDLINE-96154068; PubMed=8592061;  
RA Shimizu H., McGrath J.A., Christiano A.M., Nishikawa T., Uitto J.;  
RT "Molecular basis of recessive dystrophic epidermolysis bullosa:  
RT genotype/phenotype correlation in a case of moderate clinical  
RT severity.";  
RL J. Invest. Dermatol. 106:119-124(1996).  
RN [15]  
RP VARIANT RDEB ARG-1782.  
RX MEDLINE-96183562; PubMed=8618018;  
RA Christiano A.M., McGrath J.A., Uitto J.;  
RT "Influence of the second COL7A1 mutation in determining the  
RT phenotypic severity of recessive dystrophic epidermolysis bullosa.";  
RL J. Invest. Dermatol. 106:766-770(1996).  
RN [16]  
RP VARIANT RDEB ASP-2073.  
RX MEDLINE-96310789; PubMed=8757758;  
RA Dunnill M.G.S., McGrath J.A., Richards A.J., Christiano A.M.,  
RA Uitto J., Pope F.M., Eady R.A.J.;  
RT "Clinicopathological correlations of compound heterozygous COL7A1  
RT mutations in recessive dystrophic epidermolysis bullosa.";  
RL J. Invest. Dermatol. 107:171-177(1996).  
RN [17]  
RP VARIANTS RDEB W-1982; G-2008; A-2025; E-2049; G-2063; W-2063 AND  
RP R-2575.  
RX MEDLINE-97465605; PubMed=9326325;  
RA Hovnanian A., Rochat A., Bodemer C., Petit E., Rivers C.A., Prost C.,  
RA Fratag S., Christiano A.M., Uitto J., Lathrop M., Barrandon Y.,  
RA de Prost Y.;  
RT "Characterization of 18 new mutations in COL7A1 in recessive  
RT dystrophic epidermolysis bullosa provides evidence for distinct  
RT molecular mechanisms underlying defective anchoring fibril  
RT formation.";  
RL Am. J. Hum. Genet. 61:599-610(1997).  
RN [18]  
RP VARIANT RDEB ARG-1652.  
RX MEDLINE-98106792; PubMed=9444387;  
RA Cserhalmi-Friedman P.B., Karpati S., Horvath A., Christiano A.M.;  
RT "Identification of a glycine substitution and a splice site mutation  
RT in the type VII collagen gene in a proband with mitis recessive  
RT dystrophic epidermolysis bullosa.";  
RL Arch. Dermatol. Res. 289:640-645(1997).  
RN [19]  
RP VARIANTS DEB ARG-2009 AND ARG-2043.  
RX MEDLINE-97358588; PubMed=9215684;  
RA Winberg J.-O., Hammami-Hausali N., Nilsson O., Anton-Lamprecht I.,  
RA Naylor S.L., Kerbacher K., Zimmermann M., Krajci P.,  
RA Gedde-Dahl T. Jr., Bruckner-Tuderman L.;  
RT "Modulation of disease severity of dystrophic epidermolysis bullosa by  
RT a splice site mutation in combination with a missense mutation in the  
RT COL7A1 gene.";  
RL Hum. Mol. Genet. 6:1125-1135(1997).  
RN [20]  
RP VARIANTS DDEB ASP-1519; ASP-2006; GLU-2015 AND ARG-2034.  
RX MEDLINE-98334662; PubMed=9668111;  
RA Hammami-Hausali N., Schumann H., Raghunath M., Kilgus O., Luethi U.,  
RA Luger T., Bruckner-Tuderman L.;  
RT "Some, but not all, glycine substitution mutations in COL7A1 result in  
RT intracellular accumulation of collagen VII, loss of anchoring  
RT fibrils, and skin blistering.";  
RL J. Biol. Chem. 273:19228-19234(1998).  
RN [21]  
RP VARIANTS DEB CYS-2008; ARG-2207 AND SER-2775.  
RX MEDLINE-98410969; PubMed=9740253;  
RA Kon A., Pulkkinen L., Ishida-Yamamoto A., Hashimoto I., Uitto J.;  
RT "Novel COL7A1 mutations in dystrophic forms of epidermolysis  
RT bullosa.";  
RL J. Invest. Dermatol. 111:534-537(1998).  
RN [22]  
RP VARIANT RDEB ARG-1347.  
RX MEDLINE-99019477; PubMed=9804332;  
RA Terracina M., Posteraro P., Schubert M., Sonego G., Atzori F.,  
RA Zamburo G., Bruckner-Tuderman L., Castiglia D.;  
RT "Compound heterozygosity for a recessive glycine substitution and a  
RT splice site mutation in the COL7A1 gene causes an unusually mild form  
RT of localized recessive dystrophic epidermolysis bullosa.";  
RL J. Invest. Dermatol. 111:744-750(1998).  
RN [23]  
RP VARIANTS DEB TRP-2034; VAL-2040; ARG-2043; ARG-2064 AND ASP-2713.  
RX MEDLINE-99072663; PubMed=9856843;  
RA alignment\_scores:  
Quality: 163.00 Length: 433  
Ratio: 0.827 Gaps: 23  
Percent Similarity: 45.497 Percent Identity: 25.635  
alignment\_block:  
US-09-462-480-1 x CA17\_HUMAN ..  
Align seg 1/1 to: CA17\_HUMAN from: 1 to: 2944

821 CTTTCGACCCGCTAATACGAANAAGCGGAGCAAAAACATGCACAGACA 870  
1733 ysGlyAspProGlyLeu.....ProGlyAlaProGlyGluAArgGly 1746  
871 GCAGTGAATTCGGCGGTATCGAGC..... 897  
1747 IleGluGlyPheArgGlyProProGlyProGlnGlyAspProGlyValAr 1763  
898 .....CCGGCAAGCCGAATCCA.....GGGA 919  
1763 gGlyProAlaGlyGluLysGlyAspArgGlyProProGlyLeuAspGlyA 1780  
920 AATGTCAGTCCTCATTCCTCCTTTCACAGCGGGAAGCAGTCCTC... 966  
1780 rGserGlyLeuAspGlyLysProGlyAlaAlaGlyProSerGlyProAsn 1796  
967 .....GACCAAGCTCGCAGCGCCCTG..... 987  
1797 GlyAlaAlaGlyLysAlaGlyAspProGlyArgAspGlyLeuProGlyLe 1813  
988 .....GGCGGTACCGTTCGGAGGCGTACCAG 1015  
1813 uArgGlyGluGlnGlyLeuProGlyProSerGlyProProGlyLeuProG 1830  
1016 GGTGTCACAGAAAATGGAGCCGACCGCTACCGAGCTGAACAA...CGC 1062  
1830 lYlysProGlyGluAspGlyLysProGlyLeuAsnGlyLysAsnGlyGlu 1846  
1063 GCTGCAGAACCTGGCGCGGAGCATCAGCGAAGCCGGTCAGGCAAT 1107  
1847 ProGlyAspProGlyGluAspGlyArgGlyGlyGlyGlyGlyGlyGly 1861



OM of: US-09-462-480-1 to: SPSTREMBL\_19:\* out\_format : pfs  
Date: Jul 22, 2002 1:39 AM  
About: Results were produced by the GenCore software, version 4.5.  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:  
-MODEL=frame+n2p.model -DEV=xlh  
-O=/cgn2.1/USPTO.spool/US09462480/runat\_18072002\_164419\_19544/app\_query.fasta\_1.2850  
-DB=SPSTREMBL\_19 -QFMT=fastan -SUFFIX=rspt -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-GAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEX=7.000 -START=1 -MATRIX=blotsum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORM-ext -HEARSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09462480 -CGN1\_1\_133 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

Search information block:  
Query: US-09-462-480-1  
Query length: 1277  
Database: SPSTREMBL\_19:\*  
Database sequences: 562222  
Database length: 172994929  
Search time (sec): 143.350000

score_list:	Sequence	Strid	Orig	zScore	Escore	Len	Documentation
	sp_bacteriap:069738	+	713.00	869.06	1.3e-40	368	! O69738 mycobacterium tuberculosis
	sp_plant:Q9W08	-	210.50	246.19	2.6e-06	891	! Q92W08 arabidopsis thaliana (md
	sp_plant:Q9YFB2	+	210.50	246.16	2.6e-06	894	! Q9YFB2 arabidopsis thaliana (md
	sp_human:Q9UQ39	+	203.50	236.98	7.8e-06	956	! Q9UQ39 homo sapiens (human). r
	sp_human:Q9UQ40	+	203.50	234.42	8.2e-06	1262	! Q9UQ40 homo sapiens (human). r
	sp_human:Q9UHA8	+	203.50	228.91	9.1e-06	2296	! Q9UHA8 homo sapiens (human). s
	sp_human:Q9UQ35	+	203.50	227.24	9.4e-06	2752	! Q9UQ35 homo sapiens (human). r
	sp_bacteriap:Q33085	+	198.00	240.87	1.5e-05	302	! Q33085 mycobacterium leprae. hy
	sp_human:Q15038	+	195.50	221.45	3.0e-05	1783	! Q15038 homo sapiens (human). k
	sp_human:Q60382	+	195.50	221.41	3.1e-05	1791	! Q60382 homo sapiens (human). k
	sp_human:Q9H5F9	-	193.00	226.95	3.8e-05	704	! Q9H5F9 homo sapiens (human). c
	sp_bacteria:Q9XDH2	-	193.00	226.21	3.9e-05	763	! Q9XDH2 mycobacterium tuberculosis
	sp_rodent:Q70495	-	192.00	223.50	4.7e-05	897	! Q70495 mus musculus (mouse). pl
	sp_human:Q9UQ36	+	189.00	216.59	8.0e-05	1275	! Q9UQ36 homo sapiens (human). r
	sp_human:Q9Y2J6	-	185.00	216.32	0.0001	772	! Q9Y2J6 homo sapiens (human). k
	sp_human:Q9ULT1	-	184.50	212.25	0.0002	1124	! Q9ULT1 homo sapiens (human). k
	sp_virus:Q9Q5K9	-	182.50	215.46	0.0002	608	! Q9Q5K9 herpesvirus papio. ntr.
	sp_human:Q60585	-	180.50	210.26	0.0003	820	! Q60585 homo sapiens (human). se
	sp_bacteria:Q9EVZ2	-	180.00	220.55	0.0002	251	! Q9EVZ2 paracoccus pantotrophus
	sp_human:Q9UQ36	-	179.50	204.97	0.0004	1275	! Q9UQ36 homo sapiens (human). r
	sp_human:Q15038	-	179.50	201.88	0.0004	1783	! Q15038 homo sapiens (human). k
	sp_human:Q60382	-	179.50	201.84	0.0004	1791	! Q60382 homo sapiens (human). k
	sp_human:Q9UQ35	-	179.50	197.88	0.0004	2752	! Q9UQ35 homo sapiens (human). r
	sp_virus:Q41972	+	179.00	209.80	0.0003	706	! Q41972 murid herpesvirus 4. hyp
	sp_virus:Q41973	+	179.00	209.53	0.0003	727	! Q41973 murid herpesvirus 4. hyp
	sp_virus:Q69340	+	178.50	199.79	0.0004	1358	! Q69340 pseudorabies virus. ord
	sp_virus:Q41971	+	177.50	207.12	0.0004	774	! Q41971 murid herpesvirus 4. hyp
	sp_virtebrate:Q90YB5	-	177.00	206.39	0.0005	784	! Q90YB5 gallus gallus (chicken). o
	sp_rodent:Q35126	-	176.50	202.05	0.0006	1175	! Q35126 mus musculus (mouse). o
	sp_rodent:P70200	-	176.50	202.05	0.0006	1175	! P70200 mus musculus (mouse). o
	sp_plant:Q96397	+	176.00	207.04	0.0005	640	! Q96397 chlamydomonas reinhardtii
	sp_human:Q96V15	-	175.00	199.02	0.0007	1338	! Q96V15 homo sapiens (human). r
	sp_human:Q96RR2	-	174.50	198.42	0.0008	1337	! Q96RR2 homo sapiens (human). r
	sp_human:Q9BZG0	-	174.50	198.41	0.0008	1338	! Q9BZG0 homo sapiens (human). r
	sp_human:Q96RG4	-	174.50	198.41	0.0008	1338	! Q96RG4 homo sapiens (human). r
	sp_human:Q96RG5	-	174.50	198.40	0.0008	1339	! Q96RG5 homo sapiens (human). r
	sp_virtebrate:Q93251	+	174.50	197.70	0.0008	1445	! Q93251 rana catesbeiana (bull
	sp_rodent:Q62775	-	174.00	207.81	0.0007	451	! Q62775 rattus norvegicus (rat).
	sp_rodent:Q920G8	-	174.00	207.14	0.0007	485	! Q920G8 rattus norvegicus (rat).
	sp_bacteria:Q53913	-	174.00	206.38	0.0007	527	! Q53913 streptomyces coriofaciens

sp_rodent:Q63870	+	174.00	190.53	0.0010	2944	! Q63870 mus musculus (mouse)
sp_invertebrate:Q97406	+	172.50	195.29	0.0011	1439	! Q97406 haliotis discus (snail)
sp_virus:Q99307	-	171.00	197.54	0.0013	924	! Q99307 epstein-barr virus (virus)
sp_bacteria:Q9K3E2	+	170.50	197.86	0.0013	835	! Q9K3E2 streptomyces coelicolor
sp_virus:Q65553	-	170.50	185.35	0.0017	3247	! Q65553 bovine herpesvirus 1
seq_name: sp_bacteriap:069738						
seq_documentation_block:						
ID	O69738	PRELIMINARY;	PRT;	368	AA.	
AC	O69738:					
DT	01-AUG-1998 (TReMBLrel. 07, Created)					
DT	01-AUG-1998 (TReMBLrel. 07, Last sequence update)					
DT	01-DEC-2001 (TReMBLrel. 19, Last annotation update)					
DE	PPE-FAMILY PROTEIN.					
GN	RV3873 OR MTU027.08.					
OS	Mycobacterium tuberculosis.					
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;					
OC	Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.					
OX	NCBI_TaxID=1773;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=H37RV;					
RX	MEDLINE=98295987; PubMed=9634230;					
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,					
RA	Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekai F.,					
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,					
RA	Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,					
RA	Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,					
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,					
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R.,					
RA	Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;					
RT	"Deciphering the biology of Mycobacterium tuberculosis from the					
RT	complete genome sequence."					
RL	Nature 393:537-544 (1998).					
DR	EMBL; AL022120; CAA17965.1; -					
DR	Tuberculist; RV3873; -					
DR	InterPro; IPR000030; PPE.					
DR	Pfam; PF00823; PPE; 1.					
KW	Complete proteome.					
SQ	SEQUENCE 368 AA; 37330 MW; D78F44095F658CA2 CRC64;					
alignment_scores:						
	Quality:	713.00	Length:	143		
	Ratio:	5.057	Gaps:	0		
Percent Similarity: 98.601 Percent Identity: 96.503						
alignment_block:						
US-09-462-480-1 x O69738 ..						
Align seg 1/1 to: O69738 from: 1 to: 368						
1	CTGCAGCAGTGACGTGTTGTTTACGCCAGTGGCGGCACCGCGCGCGG	50				
226	LeuGlnGlnValThrSerLeuPheSerGlnValGlyThrGlyGlyGly	242				
51	CAACCCAGCCAGCAGGACCGCGCAGATGGCCTGTCTCGGCACCAAGTC	100				
242	YAsnProAlaAspGluGluAlaGlnMetGlyLeuLeuGlyThrSerP	259				
101	CGCTGTGCAACCATCCGCTGGCTGGTGGATCAGGCCCGCGCGCGCGG	150				
259	roLeuSerAsnHisProLeuAlaGlyGlySerGlyProSerAlaGlyAla	275				
151	GGCCTGCTGCGCGCGGAGTGCCTACCTGCGCAGGTGGTGGTTCACCCG	200				
276	GlyLeuLeuArgAlaGluSerLeuProGlyAlaGlyGlySerLeuThrAr	292				
201	CACGCCGCTGATGCTCAGCTGATCGAAAGCCGGTTGCCCTCGGTGA	250				
292	qThrProLeuMetSerGlnLeuIleGluLysProValAlaProSerValM	309				

```
251 TGCCGGCGGCTGTGTCGGATCGTCGGTGACGGGTGGCGCGCTCCGGTG 300
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
309 etProAlaAlaAlaGlySerSerAlaThrGlyGlyAlaAlaProVal 325
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
301 GGTCCGGAGGAGTGGCGGAGGTTCGCAATCCGGCGCTCCACGAGCC 350
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
326 GlyAlaGlyAlaMetGlyGlnGlyAlaGlnSerGlyGlySerThrArgPr 342
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
351 GGGTCTGTGTCGGCGGCGGCGCTCGCGAGGAGCGTGAAGAAGACGACG 400
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
342 oglyLeuValAlaProAlaProLeuAlaGlnGluArgGluGluaspGp 359
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
401 AGGAGGACTGGGACGACGAGGAGGAGTGG 429
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
359 LuAspAspTrpAspGluGluAspAspTrp 368
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

seq\_name: sp\_plant:Q9ZW08

seq\_documentation\_block:

```
AC Q9ZW08 PRELIMINARY; PRT; 891 AA.
DT 01-MAY-1999 (TrEMBLrel: 10, Created)
DT 01-MAY-1999 (TrEMBLrel: 10, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel: 19, Last annotation update)
DE PUTATIVE PROLINE-RICH PROTEIN.
GN Ar2G29210.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., VanAken S.E., Unayam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT thaliana.";
RL Nature 402:761-768(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Lin X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC004561; AAC95214.1; -.
DR InterPro; IPR002483; PWI.
DR InterPro; IPR002965; P-rich_extensn.
DR Pfam; PF01480; PWI; 1.
DR PRINTS; PR01217; PRICHEXTENS.
DR SMART; SM00311; PWI; 1.
SQ SEQUENCE 891 AA; 102261 MW; C0CCB6DE05E528B CRC64;
```

alignment\_scores:

```
Quality: 210.50 Length: 380
Ratio: 1.126 Gaps: 23
Percent Similarity: 49.211 Percent Identity: 31.316
```

alignment\_block:

```
US-09-462-480-1/rev x Q9ZW08
Align seg 1/1 to: Q9ZW08 from: 1 to: 891
1253 TCTCAGAGTGGCTCAACAGCTATATAACACGAGAAAGGCGAGACCGAGC 1204
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
298 SerArgSerProIleArgArg.....HisArgArgProThrHisGluGl 312
```

```
1203 AAGTCGAACTCGCCCGATCCCGTGTTCGGTATTCTACGGAACTCGCG 1154
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
312 yArgArgGlnSerProAlaPro.....SerArgArgArg 324
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1153 GTTCCCTATCGGAACATCCAGTACGCTTCCCTTCGGTCGACGACCATG 1104
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
324 rgSerProSerProProAlaArgArg.....ArgSerProSerPro 338
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1103 CCTCAGCGCTTCGTCGTCGCGCGAGTTCCTGCAGCGCGTGTTC 1054
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
339 ProAlaArgArgArgSerProSerProProAlaArgArgHisArgSe 355
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1053 AGTCGGTAGCGGCTCCCATTTTGTGGACACCTGGTACGCTC 1004
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
355 rProThrProProAlaArgGln.....ArgArgSerProSerProPro 370
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1003 CGAACCGCTACCGC...CCAGGCGCTGCGAGCTTGTGTCAGGGACTGCT 957
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
370 laArgArgHisArgSerProProAlaArgArgArgArgSerProSer 386
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
956 TCCCTCGTCGAGGAGGAATGAATGGACGTGACATTTCCCTGGATTGCG 907
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
387 ProProAlaArgArgArgSerProSerProProAlaArgArgArgAr 403
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
906 CTTCGCGCGCTCGATACCCGCGAAATCCACTGCTGCTCTGTCAATTT 857
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
403 gSerProSerProLeuTyArgArgAsnArgSerPro..SerProLeuTy 419
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
856 TTTGCTCGTTTCITTCGTATTAGCGGTGAGAGCCCATTTGCGAGGA 807
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
419 rArgArgAsn.....ArgSerArgSerProLeuAlaLys 431
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
806 CAGCGCTGCTGCTGCTCT...CGTCGCGCTCGAGTATTGAGCGCGG 760
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
431 rgGlyArgSerAspSerProGlyArgSerProSerProValaArgLeu 447
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
759 CCTGACGAATATTCGTCGAGA.....TCTGTCGAGT 728
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
448 ArgAspProThrGlyAlaArgLeuProSerProSerIleGluGlnArgLe 464
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
727 TCCGTGCTCTGCTATTGGCTGCTTCTTGGAGCGCACCCG...CGCG 681
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
464 uProSerProProValaGlnArgLeuProSerProProProArgArg 481
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
680 CTGGCGCGCTGCCCGCGCGCGCCACTGCGCTGCAACGAACTG 631
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
481 laGlyLeuProSerProPro.....ProAlaGlnArgLeu 492
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
630 CCGTCGACTCCACTGTCGATCTGGGTTTTCAGGTGCGCGGAGA..... 586
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
493 ProSerProProArgArgAlaGlyLeuProSerProMetArgIleGl 509
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
585 .....TCCGCTCGAAATTACCTG..... 568
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
509 yGlySerHisAlaAlaAsnHisLeuGluSerProSerProSerSerLeu 526
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
567 ..CCTCCTGCCCGAGGTAGCGCATCGGCTTCATCTCTGCCATGCTGG 520
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
526 erProProGlyArg.....LysLysValLeuProSerProProValArg 540
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
519 ACTACTTTCTCTTTTACCTTCCTCGCCAAATTTGGCAAGTCTTCGG 470
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
541 ArgArgArgSerLeuThrProAspGluGluArgValSerLeuSerGln.. 556
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
469 CCGGGTGGCGGGAAGTCTGTTGTCATTAC.....GGGAGC. 433
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
557 ....GlyGlyArgHisThrSerProSerHisIleLysGlnAspGlySer 572
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
432 ..TCACAGTCTGCTCTCGTCCAGTCGTCCTGCTCTCTCTCTTCACG 385
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
572 etSerProVal.ArgGlyArgGlyLysSerProSerSerArg.HLsg 588
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
384 CTCTGCGGAGCGGTGCCGCGCGACGACCG.....GGCTGG 344
```

```
|||||... ||| ...|||
588 lnlysAlaArgSerProValArgArgSerProThrProValAsnArg 604
ID 09FVB2 PRELIMINARY; PRT; 894 AA.
AC Q9FVB2;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE SRM102.
GN SRM102.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae;
OC eucotyledons; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID:3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Lopato S., Weber K., Kalyana M., Meissner M., Langhammer U., Barta A.;
RT "AtRSP31 functions in a complex with a hSRM160-like protein.";
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ292982; CAC03679.1;
DR InterPro; IPR002483; PWI.
DR InterPro; IPR002965; P-rich_extensn.
DR Pfam; PF01480; PWI; 1.
DR PRINTS; PR01217; PRICHEXTENS.
DR SMART; SM00311; PWI; 1.
SQ SEQUENCE 894 AA; 102378 MW; 705C5DC2527A6C75 CRC64;
```

```
alignment_scores:
  Quality: 210.50      Length: 380
  Ratio: 1.126         Gaps: 23
  Percent Similarity: 49.211 Percent Identity: 31.316

alignment_block:
US-09-462-480-1/rev x Q9FVB2 ..
Align seg 1/1 to: Q9FVB2 from: 1 to: 894

1253 TCTCAGAGTCGGCTCAACAGCTATAACACGAGAAAGGCGGAGACCGG 1204
|||||... ||| ...|||
301 SerArgSerProIleArgArg.....HisArgArgProThrHisGluG 315
1203 AAGTTCGAAGTCCCGGATCCGCTGTTGGTATCTACGGAACTCGG 1154
|||||... ||| ...|||
315 yArgArgGlnSerProAlaPro.....SerArgArgArg 327
1153 GTTCCCTATCGAAGATCCAGTACGTTGCTTGGTCCGTCGAGCCATTG 1104
|||||... ||| ...|||
327 IgSerProSerProAlaArgArg.....ArgSerProSerPro 341
1103 CCTGACCGGCTCGCTGATCGTCGCGCCAGGTTCTGCAGCGCTGTTCT 1054
|||||... ||| ...|||
342 ProAlaArgArgArgSerProSerProProAlaArgHisArgSe 358
1053 AGTCGGTAGCGGGGTCCCATTTTGTGGACACCCCTGGTACGGCTC 1004
|||||... ||| ...|||
358 rProThrProProAlaArgGln.....ArgArgSerProSerPro 373
```

```
1003 CGAACCGCTACCG...CCGAGCGCGCTGGAGCTTGGTCAGGACTGCT 957
|||||... ||| ...|||
373 laArgArgHisArgSerProProAlaArgArgArgArgSerProSer 389
|||||... ||| ...|||
956 TCCCTCTGTCGAAGGAGGAATGAATGACGTCGACATTTCCCTGGATTGCG 907
|||||... ||| ...|||
390 ProProAlaArgArgArgArgSerProProProAlaArgArgArgAr 406
|||||... ||| ...|||
906 CTTCGCGCGGCTCGATACCGCGGAATTCACATGCTGCTCTCTCATGTT 857
|||||... ||| ...|||
406 gSerProSerProLeuTyrArgArgAsnArgSerPro..SerProLeu 422
|||||... ||| ...|||
856 TTTGCTCCGTTCTTTTCGATTAGCGGTCAGAAGCCCATTTGCGAGGA 807
|||||... ||| ...|||
422 rArgArgAsn.....ArgSerArgSerProLeuAlaLysA 434
|||||... ||| ...|||
806 CAGCGCTGCTGCTGCTCT...CGTCGCGCTCGAGTATTGGACGCGG 760
|||||... ||| ...|||
434 rgGlyArgSerAspSerProGlyArgSerProSerProProAlaArgLeu 450
|||||... ||| ...|||
759 CTGACGAATATTCGTCGAGA.....TCTCGTCGAGT 728
|||||... ||| ...|||
451 ArgAspProThrGlyAlaArgLeuProSerProSerIleGluGlnArgLe 467
|||||... ||| ...|||
727 TCTGCTGCTGCTTATTGGCTGCTTCTTGAAGCGCACACCG...CGGC 681
|||||... ||| ...|||
467 uProSerProProValAlaGlnArgLeuProSerProProProArgArgA 484
|||||... ||| ...|||
680 CTGGCGGCGCTCCCGCGCGCGCGCCCTGCGCCCTGCAACGAACTG 631
|||||... ||| ...|||
484 laGlyLeuProSerProPro.....ProAlaGlnArgLeu 495
|||||... ||| ...|||
630 CGTCGACTCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 586
|||||... ||| ...|||
496 ProSerProProProArgArgAlaGlyLeuProSerProMetArgIleG 512
|||||... ||| ...|||
585 .....TCCGCTCGAAATACCTG..... 568
|||||... ||| ...|||
512 yGlySerHisAlaAlaAsnHisLeuGluSerProSerProSerSerLeu 529
|||||... ||| ...|||
567 ..CCTCTCGCGCGAGGTAGCGGCTGCTTTCATCTCTGCTGCTGCTGCT 520
|||||... ||| ...|||
529 erProProGlyArg.....LysLysValLeuProProProValArg 543
|||||... ||| ...|||
519 ACTACTTCTCTCTTACCTTCTCTCGCCAAATGTTGGCAAGTCTTCGG 470
|||||... ||| ...|||
544 ArgArgArgSerLeuThrProAspGluArgValSerLeuSerGln... 559
|||||... ||| ...|||
469 CCGGCTGCGCGGGAAGTCTGTTGTCATTAC.....GGGAGC. 433
|||||... ||| ...|||
560 ....GlyGlyArgHisThrSerProSerHisIleLysGlnAspGlySerM 575
|||||... ||| ...|||
432 ..TCACAGTCGCTCTCTGTCCTCCAGTCGCTGCTGCTGCTGCTGCTC 385
|||||... ||| ...|||
575 etSerProVal.ArgGlyArgGlyLysSerSerProSerSerArg.HisG 591
|||||... ||| ...|||
384 CTCCTGCGCGAGCGGTGCGCGCGGACGACCGG.....GGCTGG 344
|||||... ||| ...|||
591 lnLysAlaArgSerProValArgArgSerProThrProValAsnArg 607
|||||... ||| ...|||
343 TGGAGCGCGCGGATTGCGAACCCCTGCGCCCTGCTCCCGCACCCCGGA 294
|||||... ||| ...|||
608 ArgSerArgArgSerSerSerAlaSerArgSer...ProAspArgArgAr 623
|||||... ||| ...|||
293 GCGCGCGCACCGG.....TACCCGACGATCCGGCAACAGCGCG 256
|||||... ||| ...|||
623 gArgArgSerProSerSerArgSerProSerArgSerArgSerSerProp 640
|||||... ||| ...|||
255 CGGCA.....TCACCGAGGGGCAA 236
|||||... ||| ...|||
640 roValLeuHisArgSerProSerProArgGlyArg 651
```







DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE RNA BINDING PROTEIN.  
OS Homo sapiens (human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Ohtaki S., Umeki K., Sawada Y.;  
RT "Homo sapiens mRNA for RNA binding protein, complete cds.";  
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB016092; BAA83718.1;  
DR InterPro; IPR002965; P\_Rich\_extensn.  
DR PRINTS; PR01217; PRICHTEXTENS.  
SQ SEQUENCE 2752 AA; 299672 MW; 109C64F181097123 CRC64;

## alignment\_scores:

Quality: 203.50 Length: 442  
Ratio: 0.988 Gaps: 27  
Percent Similarity: 46.606 Percent Identity: 28.733

## alignment\_block:

US-09-462-480-1 x Q9UQ35

Align seg 1/1 to: Q9UQ35 from: 1 to: 2752

33 GGGCGGACCGCGGGCGGACACCGCGGACGAGA...AGCCGCGGAGA 79  
||| |||||::: :::: ||| :::: |||||:::  
448 GlySerHisArgGluIleSerSerSerProThrSerLysAsnArgSerHi 464  
80 TGGCGCTGCTGGCAGCAGTCCGCTGTCGACCATCCCTGGCTGGTGA 129  
||| |||:::||||:::||||::: |||||  
464 sGlyArgAlaLysArgAspLysSer.HisSerHisThr..... 476  
130 TCAGGCCCGACCGCGGGCGGCGGTGCTGCGCGGAGTCGCTACCTGG 179  
||||| ||| |||||::: |||::: |||  
477 .....ProSerArgArgMetGlyArgSerArgSerProAlaThrAlaLy 491  
180 CGCAGGTGGTGGTGTGACCCCGACCGCGCTGATCTCAGCTGATCGAA 229  
||| |||:::|||||||::: |||::: |||:::  
491 sArgGlyArgSerArgSerArgThrProThrLysArg.GlyHisSerArg 507  
230 AGCCGGTGGCCCTCGGTGATCGCGGCGCTGTCGCGGATCGTCGGTG 279  
||||| |||  
508 SerArgSerPro..... 511  
280 ACGGTGGCGCGCTCCGTTGGTGGTCCGG.....GAGCGATGGG 317  
::: ||| ||| |||||::: |||::: |||  
512 .....GlnTrpArgArgSerArgSerArgAlaGlnArgTrpG 523  
318 CCAGGTGGTGGCAATCCG.....GCGGTCCACCAACC 349  
::: ||| |||::: |||::: |||::: |||  
523 LysArgSerArgSerProGlnArgGlyArgSerArgSerProGlnArg 539  
350 CGGTCTGCTGGCGCGGCGCGCTCGCGCAGGAGCGGTGAGACAGCAC 399  
||||| |||  
540 ProGlyTrpSer.....ArgSerArgAsnThrGln.ArgArgGlyA 553  
400 GAGGACGACTGGGACGAGGAGGAGGAGTGGTGAGTCCCGTAATGACAA 449  
|||::: |||::: |||::: |||::: |||  
553 rGlyArgSerAlaArgGly.....ArgSerHisSer 564  
450 AGACTTCCCGCCCGCGCGCGGCGGAGAGACTTCCCAACATTTTGGCAGGA 499  
||| ||||| ||||| |||||  
565 Arg.SerProAlaThrArg...GlyArg..... 572  
500 AGGTAACAGAGAAAGTAGTCCAGATGGCAGATGAAGACCGATGCCG 549  
::: |||::: |||::: |||  
573 .....SerArgSerArgThrProAla..... 579

550 CTACCTCGGCGAGGAGGAGGAGTAATTTTCAGCGGATCTCCGGCGACCTG 599  
||||| ||| :::: |||||  
580 .....ArgArgGlyArgSerArgSerArgThrProAlaArgAr 592  
500 AAAACCCAGATCGACAGGAGGAGTGGAGTGCAGCGCAGGTTCTGTCAGGCGCA 649  
::: |||||::: |||||::: ||| |||:::  
592 gArgSerArgSerArgThrProThrArgArgSerArgSerArgThr. 608  
650 GTGGCGCGGCGGCGGCGGCGGCCCGCCAGCGCGGTGCTGCTGCTCC 699  
||||| ||||| |||||  
609 .....ProAlaArgArgGlyArg.....Ser 615  
700 AAGAACGAGCCCAATAGCAGAGCAGGAACCTCGACGATCTCGACGAAT 749  
::: |||||::: |||||::: |||||  
616 ArgSerArgThrProAlaArgArgSerArgThrArgSerPro..... 630  
750 ATTGCTCAGCGCGGCTGCCAATACTCGAGGCGCCAGCAGGAGCAGCAGCA 799  
||||| |||||::: |||||::: ||| |||:::  
631 ...ValArgArgArgSerArgSerArgSerProAlaArgSerGlyA 646  
800 GCGCTGCTCTCGCAATGGCTTCTGACCGGCTAATACGAAAAAGAACG 849  
||| |||::: |||  
646 rGSer.....ArgSerArgThr 651  
850 GAGCAAAAACATGACAGCAGCAGTGGAAATTCGCGGATCGAGCGCG 899  
::: |||||::: |||::: |||  
652 ProAlaArgArgGlyArgSerArgSerArgThrProAlaArgArgGlyAr 668  
900 CGGCAAG...CGCAATCCAGGGAATGTCAGCTCCATTCCTCCTT 946  
|||::: |||::: |||::: |||  
668 gSerArgSerArgThrProAlaArg..... 676  
947 GACGAGGGAAGCAGTCCCTGACCAAGCTCGCAGCGCTCGGCGGTAG 996  
||||| |||||::: |||  
677 .....ArgSerGlyArgSerArgSer 683  
997 CGGTTCGAGGCGTACCAGGCGTCCAGCAAAATGGACGCCACGCGCTA 1046  
||| |||||::: |||||  
684 Arg.....ThrProAlaArgArgGlyArgSerArgSe 694  
1047 CCGAGCTGACACACGCGCT..... 1065  
|||::: |||::: |||  
694 rArgThrProArgArgGlyArgSerArgSerArgSerLeuValArgArgG 711  
1066 .....GCAGAACCTGGCGGCGGACGATCAGCAAGCGGTCAGCAATGGC 1110  
711 LysArgSerHisSerArgThrProGlnArgArgGlyArgSerGlySer 727  
1111 TTCACCGAAGCAACGTCACCTGCTGATGTCGATAGGCGCAACGCCAGT 1160  
::: |||||::: ||| |||||  
728 SerGluArgLysAsnLysSer.....ArgThrSerGlnArgArgSe 741  
1161 TCGCGTGAATAGCGAA.....ACACGGATCGGCGGAGTTCGACC 1201  
||| |||||::: |||||  
741 rArgSer.AsnSerSerProGluMetLysLysSerArgIleSerSer... 756  
1202 TTCGTCGCTCGCCCTTCTCGTGTATACGTTTCAGCGCAGCTCTGA 1251  
||||| |||||::: ||| |||  
757 ...ArgArgSerArgSerLeuSerSerProArgSerLysAlaLysSer... 771  
1252 GAGGTGTGTCATGGCGG 1267  
||||| |||  
772 ArgLeuSerLeuArg 776  
seq\_name: sp\_bacteriap.033085  
seq\_documentation\_block:  
ID 033085 PRELIMINARY; PRT; 302 AA.  
AC 033085;  
DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE HYPOTHETICAL 32.1 KDA PROTEIN (PPE-FAMILY PROTEIN).





```

:::||||:::||||
871 Gly ArgSerArgThrProProThrSerArgLysArgSerArgSerArg 887
1181 .....CGGACGCGGAGCT...TCGACC 1201
      |||:::|||||||:::||||
887 hrSerProIaProTrpLysArgSerArgSerArgAlaSerProAlaThr 903
1202 TTCGTCGCTCGC 1216
      ||||||||||
904 HisArgSerArg 908
seq_name: sp_human:O60382
seq_documentation_block:
ID O60382 PRELIMINARY; PRT; 1791 AA.
OS O60382;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE KIAA0324 (FRAGMENT).
GN KIAA0324.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleo-
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ricke D.O., Bruce D., Mundt M., Doggett N., Munk C., Saundel-
RA Robinson D., Jones M., Buckingham J., Chasteen L., Thompson
RA Goodwin L., Bryant J., Tesmer J., Meincke L., Longmire J.,
RA Ueng S., Tatum O., Campbell C., Fawcett J., Deaven L.;
RT "Sequencing of Human Chromosome 16p13.3";
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Ricke D.O.;
RT "Large Scale Sequence Analysis and Annotation with the Seq-
RL Comparison Analysis (SCAN) System.";
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC004493; AAC08453.1; -.
FT NON_TER 1
SQ SEQUENCE 1791 AA; 191306 MW; 3A7B5530AEP95F3E CRC64;

alignment_scores:
Quality: 195.50 Length: 474
Ratio: 0.846 Gaps: 24
Percent Similarity: 48.734 Percent Identity: 27.215

alignment_block:
US-09-462-480-1 x O60382 ..

Align seg 1/1 to: O60382 from: 1 to: 1791

48 CGGCAACCCAGCGCAGGAGCGCGCAGATGGGCT.....GC 88
||| ||| ||| ||| ||| ||| ::::: |||
465 ArgThrProSerArgArgSerArgSerGlySerProGlyLeuArgAs 481
89 TCGGCACCAAGTCGCTGTCGAACATCCGCTGGCTGGTGGATCA..... 132
|||::: |||::: |||::: |||::: |||::: |||
481 p.GlySerGlyThrProSerArgHisSerLeuSerGlySerProGly 497
133 .....GGCCCCAGCGCGCGCGGCTGCTGGCGGC 164
||| ||| ||| ||| ||| |||
498 MetLysAspIleProArgThrProSerArgGly.....ArgSe 510
165 GGAG .....TCGCTACCTGGCGCAGGTGGGT 190
|||
510 rGlucyAspSerSerProGluProLysAlaLeuProGlnThrProArgP 527
191 CGTTGACCCGCGCGCGTGATCTCTAGCTGATCGAAAAGCCGTTGCC 240
::: |||::: |||
527 roArgSerArgSerProSerSerProGluLeuAspAsnLysCvsLeuThr 543

```

```
241 CCCTCGGTGATCCGGCGGCTGTTCCGGATCGTCGCTGAGCGGTGGCGC 290
|||||
544 ProGln.....ArgGluArgSerGlySerGluSerValAspGI 557
291 CGCTCCGGTGGTCCGGGAGGATGGCCAGCGGTTCGCAATCCGGCGCT 340
|||||
557 nLysThrValAlaArgThrProLeuGlyGlnArgSerArgSerGlySers 574
341 CCACAGCCCGGCTGCTGGTCGCGCGGCGCACCGCTCGCGCAGGAGCGTAA 390
|||
574 erGlnGluLeuAsp...ValLysProSerAlaSerProGlnGluArgSer 589
391 GAAGCAGCAGGACGACTGGGACGACGAGCAGCAGCTGGTGAGCTCCCGT 440
|||||
590 GluSer.....AspSerSerProAspSerLysAlaLysTh 601
441 AATGACAACACACTCCCGCGCCACCGCGGCGGAGAGACTTCCCAACATT. 489
: |||
601 rArgThrProLeuArgGlnArgSerArgSerGlySerSerProGluValA 618
490 .....TTGCGAGGAAGGTAAAGAGAGAAAGTAGTCCAGCA 525
|||||
618 spSerLysSerArgLeuSerProArgArgSerArgSerGlySerSerPro 634
526 TGCAGAGATGAACCCGATCCGCTACCCCTCGGCGCAGGAGGAGTAAT 575
: |||
635 GluValLysAspLysProArgAlaAlaProArgAlaGlnSerGlySerAs 651
576 TTCAGCGGATCTCCGGGACCTGAACACCAGATCG...ACCAGTGGGA 622
|||||
651 pSerSerProGluProLysAlaProAlaProArgAlaLeuProArgS 668
623 GTCAGCGCAGGTTCTGTCAGGCGCAGTGGCGGCGCGCGGCGGAGCG 672
|||||
668 erArg.....SerGlySerSerLysGlyArg 677
673 CCGCCCGCGCGGTGGTGGCTTCCAGAAAG.....CAGCCAAAT 713
|||
678 GlyProSerProGluGlySerSerSerThrGluSerSerProGluHisPr 694
714 AAGCAGAGCAGGAACTCGCAGCAGATCTCGACGAATATTCGTACGGCGG 763
: |||
694 oProLysSerArgThrAlaArgArgGlySerArgSerSerProGluProL 711
764 CGTCCAATACTCGA...GGGCGCAGCAGGAGCAGCAGCAGCGCTGCTCT 810
|||||
711 ysThrLysSerArgThrProProArgArgArgSerSerArgSerSerPro 727
811 .....CGCAAAATGGGCTTC 824
728 GluLeuThrArgLysAlaArgLeuSerArgArgSerArgSerAlaSerSe 744
825 TGACCCGCTAATAGCAAAAGAACGGAGCAAAACATGACAGACGACAG 874
|||||
744 rSerProGluThrArgSerArgThrProProArgHisArg...SerPr 760
875 TGGAAAT.....TCGCGGATATCGAGCGCGC 900
: |||
760 oSerValSerSerProGluProAlaGluLysSerArgSerArgArgA 777
901 GGCAAGCGCAATCCAGGAAATGTCAGCTCATTCATTCCTCCTCCTCAGC 950
|||||
777 rArgSerAlaSerSerProArgThrLys.....ThrThrSerArg 790
951 AGGGAGAGCAGTCCCTGACCAAGC..... 974
791 ArgGlyArgSerProSerProLysProArgGlyLeuGlnArgSerArgSe 807
975 ....TCCAGCGGCTCGGCGGTTAGCGGTTCGAGCGGTACACAGGTGT 1020
|||||
807 rArgSerArgGluGlySerArgThrThrArgArgArgArgArgSerg 824
```

```
1021 CCAGCAAAATGGACGCCACGCTACCGAGCTGAACAACGCGTGCAGA 1070
: |||
824 lySerSerGlnSerThrSerArg.ArgArg.....GlnArgSerArgSe 838
1071 ACCTGGCGCGGACGATCAGCGAAGCCGCTCAGGCAATGGCTTCGACCGAA 1120
: |||
838 rArgSerArgValThrArgArgArgArgGlyGlySerGlyTyrHisSerA 855
1121 GGCAAA.....CGTCACTGGGATGTTCGATAGGCAACGCCGAGTTCGC 1164
|||||
855 rArgSerProAlaArgGlnGluSerSerArgThrSerSerArgArgArg 871
1165 GTAGATAGCGAAACA..... 1180
872 Gly.ArgSerArgThrProProThrSerArgLysArgSerArgSerArgT 888
1181 .....CGGGATCGGCGGAGT...TCGACC 1201
888 hrSerProAlaProTriLysArgSerArgSerArgAlaSerProAlaThr 904
1202 TTCGTCGCTCGCTCGC 1216
905 HisArgArgSerArg 909
seq_name: sp_human:Q9H5F9
seq_documentation_block:
ID Q9H5F9 PRELIMINARY; PRT; 704 AA.
AC Q9H5F9;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CDNA: FLJ23471 FIS, CLONE HS111969.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SMALL INTESTINE;
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK027124; BAB15667.1;
DR InterPro: IPR002965; P_rich_extensn.
DR PRINTS: PRO1217; PRICHEXTENS.
SQ SEQUENCE 704 AA; 75768 MW; 858F94EEA2C1F8C6 CRC64;
```

## alignment\_scores:

Quality: 193.00 Length: 504  
Ratio: 0.881 Gaps: 30  
Percent Similarity: 43.452 Percent Identity: 25.992

## alignment\_block:

US-09-462-480-1/rev x Q9H5F9 ..

Align seg 1/1 to: Q9H5F9 from: 1 to: 704

```
1217 GCGCAGACCGCAGGAGGTGCAAC.....TCGCCGATCCCGTGT 1177
92 GlyAsnSerProAlaArgAlaSerValProAlaAlaProAsnProAlaAl 108
1176 TCGCTATTCTACGCAACTCGG...CGTTGCCCTATGCGAACATCCAGT 1130
108 aThrSerAlaThrSerValHisValArgSerProAlaArgProSerGluS 125
1129 GAGCTTCCTTCGTCGAAGCCATTGCTCACCAGGCTTCGC..... 1089
125 erArg...LeuAlaProThrProThrGluGlyLysValArgProArgVal 140
```

1088 TGATCGTCGCGCCA.....GGTCTCAGCGGTTGTTTCTAGCTCG 1048  
:::|||||  
141 ThrAsnSerProMetGlyTrpSerSerAlaAlaProCysThrAla... 156  
:::|||||  
1047 GTAGCGTGGCGTCCATTTTGTGACACCCCTGGTACGCTCCGACCC 998  
:::|||||  
157 .....AlaAlaAlaSerHisProAlaValProSerA 168  
:::|||||  
997 GCTACCGCCCGCAGCGCTGGAGTCTGGTCTGGTCTGGTCTGGTCT 948  
|||||  
168 laProAspProArgPro.....AlaThrProGln 177  
|||||  
947 CAAGGAGGAATGAATGACCTGACATTCCTGCTGATTCGCTTCGCGC 898  
|||||  
178 GlyGlyGly.....AlaProArgValAlaAlaProG1 188  
|||||  
897 GCCTCGATACCGCGGAATTCACCTGCTGCTCTCATGTTTTCCTCCG 848  
:::|||||  
188 nThrThrLeuSerSerSerThrSerAlaAlaThr.ValAspProPro 204  
:::|||||  
847 TTTCTTTTCTGATTAGCGGGTCAGAAGCCCATTTTCGAGGACA..... 805  
|||||  
205 AlaIrpThrProSerAlaSer.ArgThrGlnGlnAlaArgAsnLysPheP 221  
:::|||||  
804 ..GGCCCTGCTGCTCTCTCGTCTCGCCCTCGAGTATTGGACGCGGCT 757  
:::|||||  
221 heGlnThrSerAlaValProGlyThrSerLeuSerGlyArgGlyPro 237  
:::|||||  
756 .....GACCAATATTCGTCGAGATCTGCTCG 731  
|||||  
238 ThrProSerLeuValLeuSerLysAspSerSerLysGluGlnAlaArgAs 254  
:::|||||  
730 AGTTCTCTGCTCTGCTATTGCTGCTCTTGTGGAAGGCGACACCGCGGC 681  
:::|||||  
254 nPheLeuLysGlnAlaLeuSerAlaLeuGluAlaGlyAlaProAlap 271  
:::|||||  
680 CTGGCGCGCGCTCCCGCGCGCGCGCGCC.....ACTGGCCCTGCA 640  
|||||  
271 roGlyArgProSerProAlaThrAlaAlaValProSerSerGlnProLys 287  
:::|||||  
639 ACGAACCTGCGG.....TCGACTCC 620  
|||||  
288 ThrGluAlaProGlnAlaSerProLeuAlaLysProLeuGlnSerSerSe 304  
:::|||||  
619 ACCTGGTCTGATCTGGTTTTCAGTCTCGCGCGAGA..... 586  
|||||  
304 rProArgValLeuGlyLeuProSerArgMetGluProProAlaProLeuS 321  
:::|||||  
585 .....TCCGCTCGAATACCTGCTCTCT..... 562  
|||||  
321 erThrSerSerThrSerGlnAlaSerAlaLeuProProAlaGlyArgArg 337  
:::|||||  
561 .....GCCGAGGTACGCGCATCGGCTTCATCTCTGCCATCTGGAC 518  
|||||  
338 AsnLeuAla.GluSerSerGlyValGly.....ArgValGlyA 350  
:::|||||  
517 TACTTTCT.....CTCTTT 504  
|||||  
350 laGlySerArgProLysProGluAlaProMetAlaLysGlyLysSerThr 366  
:::|||||  
503 ACCTTCCTCGCCAAATTTGGCAGTCTT..... 474  
|||||  
367 ThrLeuThrGlnAspMetSerThrSerLeuGlnGluGlyGlnGluAspG1 383  
:::|||||  
473 ..CCGCGCGCGGTGGCGGGAAGTCTG.....TTCTCATACGGGAGCTC 431  
|||||  
383 yProAla..GlyTrpArgAlaAsnLeuLysProValAspArgArgSerPro 399  
:::|||||  
430 ACCAGTCTCTCTCTGCTCTGCTCTGCTCTCTCTCTCTCTCTCTCT 389  
:::|||||  
400 AlaGluArgThrLeuLysProLysGluProArgAlaLeuAlaGluProAr 416  
:::|||||  
388 .....CACGCTCTGCGGAGCG..... 371

416 gAlaGlyGluAlaProArgLysValSerGlySerPheAlaGlySerValH 433  
:::|||||  
370 .....GTCCGCGCGGACACCGCGGCTGGTGGAGCGCGCG 332  
:::|||||  
433 isIleThrLeuThrProValArgProAspArg.....ThrProArg 446  
:::|||||  
331 ATTGCGAACCTGCGCTCTGCTCCGAGACCCACCGGAGCGCGCCACCC 282  
|||||  
447 ProAlaSerProGlyProSerLeuPro.....AlaArgSerPr 459  
:::|||||  
281 GTCACCGACGATCCGCGCAACAGCGCGCGGCATCACCGAGGGGCAACCGG 232  
|||||  
459 oSerProProArg..... 464  
:::|||||  
231 CTTTTCGATCATCTGAGACATCATCGCGGCTCGCGGTCAACGACCCACCTG 182  
:::|||||  
464 ..... 464  
181 CGCCAGGTAGCTAGCTCGCGCGGACAGCGCGCGCGCTGGGCGCT 132  
|||||  
465 ArgArgLeuAlaValProAlaSerLeuAspValCysAspAsnTrpLeuAr 481  
:::|||||  
131 GATCCACGACCGACGCG...GATGTTTCACAGCGGACTGTGTCGCGAGCAG 85  
:::|||||  
481 gProGluProProGlyGlnGluAlaArgValGlnSerTrpLysGluGluG 498  
:::|||||  
84 GCCCATCTGCGCGGCTCTCTGCTGCGGTGGTTCGCGCGCGGTCGCGCG 35  
:::|||||  
498 luLysLysProHisLeuGlnGlyArgProGly.....Arg 509  
:::|||||  
34 CCA.....CCTGCTGAACAAGCAGCTC 12  
|||||  
510 ProLeuSerProAlaAsnValProAlaLeuProGlyGluThrValThrSe 526  
:::|||||  
11 ACCT 8  
:::|||||  
526 rPro 527

seq\_name: sp\_bacteria:Q9XDH2  
seq\_documentation\_block:  
ID Q9XDH2 PRELIMINARY; PRT; 763 AA.  
AC Q9XDH2;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE PROLINE-RICH MUCIN HOMOLOG.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=H37RV;  
RC MEDLINE=20090472; PubMed=10627046;  
RA Espitia C., Lacleite J.P., Mondragon-Palmino M., Anador A.,  
RA Campuzano J., Martens A., Singh M., Cicero R., Zhang Y., Moreno C.;  
RT "The PE-PGRS glycine-rich proteins of Mycobacterium tuberculosis: a  
RL Microbiology 145:3487-3495(1999).  
DR EMBL; AF071081; AAD41594.1; -  
DR InterPro; IPR002951; Atrophin.  
DR InterPro; IPR003882; Pistil\_extensin.  
DR PRINTS; PRO1222; ATROPHIN.  
DR PRINTS; PRO1217; PRICHEXTENSIN.  
DR PRINTS; PRO1218; PSTLEXTENSIN.  
SQ SEQUENCE 763 AA; 75035 MW; 39168EC45A5916F8 CRC64;

alignment\_scores:  
Quality: 193.00 Length: 435

Ratio: 1.078 Gaps: 21  
Percent Similarity: 41.149 Percent Identity: 27.126

## alignment\_block:

US-09-462-480-1/rev x Q9XDH2

Align seg 1/1 to: Q9XDH2 from: 1 to: 763

```
1124 TGCCTTCGTCGACGAGCCATTCCTGACGCGCTTCGCTGATGCTCGCGCC 1075
|||||
281 CysProProLeuProProLeuProAsn.....AsnHisProPr 293
1074 AGGTTCTGCACGCGGTGTTTCAGCTCGGTAGCGTGGCGTCCATTTTGG 1025
|||||
293 oAlaProProAlaProValProGlyValProLeuAlaProLeuProA 310
1024 CTGGACACCTGGTACGCTCCGACCGCTACCGCC 987
|||||
310 snSerHisProProAlaProProSerAlaProValProGlyValProLeu 326
986 .....AGGCGGTGGAGCTTGGTCAAGGACTGC 958
|||||
327 AlaProLeuProIleSerGlyArgProValSerValTrpLysGlySerPh 343
957 TTCC..... 954
343 eThrThrLeuSerThrPheCysCysArgValCysSerGlyGluValLeuA 360
953 .....CTCGTCAAGGAGGAAATGATGACGCTGACAT 921
360 laGlyAlaLeuAsnProSerArgProSerArgSerProLeuThrThr 376
920 TTCCTGATTGCGCTGCGCGCGCTCGATACCGCGGAAATTCACCTGC 871
377 ThrProAlaLeuProAlaProIleProProLeuProProLeuProProL 393
870 TGCTCTGTGATGTTTTCGCTCCGTTCTTTCGTTATAGCGGGTCAGAAG 821
|||
393 euProIleAsnThrAlaValProProIleProProLeuProProVal.Th 409
820 CCCATTTCGAGGACAGCGCTGCTGCTGCTCCTGCTCGCGCCCTCGAGTA 771
409 iAlaLeuAlaProProLeuProProLeuAlaPro...LeuProIleSerP 425
770 TTGGACGCGCGCTCGACGAATATTCGTCGAGATCTCGTCGAGTCCCTGCT 721
425 roGlyValProPro.....AlaProProIleProPro 435
720 TCTGCTTATGCGTGT.....CTTGGAGCGCACACGCGG..... 682
436 GlyLysProTrpThrThrProProLeuAlaProAlaProProGluProLy 452
682 ..... 682
452 sThrValProValLeuProProGlyProSerCysProSerGluLysP 469
681 .....CCTGGGCGCGCTCCCGCGCGCGCTGCGCTGCGCTGCA. 640
469 roAsnProProAlaProProGluProProGluProLysSerSerProAla 485
639 ACGAACTCGCGTCACTCACTGGTGGTCACTGGTGGTTCAGTTCGCGCG 590
486 .....LeuProProAlaProProAlaProSerMetProSerAlaValAr 500
589 GAGATCCGCTCGAAATACCTGCTGCTGCGCGGAGGTAGCGCATCGT 540
500 gValProProSerProIleProProAlaProProAlaProAlaProArgA 517
539 CTTTCATCTCTGCCATGCTGGACTACTTCTCTCTTTACCTTCCTCGCAA 490
517 laSer.....MetProAlaLeu 522
489 AATGTTGGCAAGTCTTCGGCGCGCGGTGGCGGGAAGTCTGTTCTCATTA 440
```

```
523 ProProAlaProProSerProProAlaThrArg...LeuCysProProLe 538
439 CGGAGGCTCACCAGT.....CGTCTCTTCTGTCCTCCA 409
538 uProProSerProProAlaProAsnSerProProAlaProProAlaProP 555
408 GTCGCTCCTGCTGCTTCTTTCACGCTCCTCGCGAGCGGTGCGCGCGCA 359
555 roThrProLysLeuLeuSerAlaAsnProProCysProProValPro 571
358 CCACACCGCGGCTGGTGGAGCGCGGATTCGGAACCTCTGGC..... 317
572 ProAlaPro.....AsnArgProProAlaProProAlaProAl 585
316 ...CCATCGCTCCGCGACCGGAGCGCGCGCACCGCTTCGATCA 271
585 aProProGluLeuProAlaProProAsP.....ProProThrProProV 600
270 TCCGGCAACAGCGCGCATCACGAGGGGCAACCGGCTTTTCGATCA 221
600 alAlaAsnSerProProAlaProProAla.....ProPro 611
220 GCTGAGACATCAGCGCGTGGGGTCAACGACCGCTGCGCGCGAGGTAGC 171
612 AlaProProSerAla..LeuProPheValAsnProProAlaProProThr 627
170 GACTCCGCG...CGCAGCAGCGCGCG...CCGCGCTGGGCGCTGATCC 127
628 ProAlaAlaProLysSerArgProAlaLeuProAlaAlaProProAlaPr 644
126 ACCAGCCAGCGGATGTTTCGACGCGGACTGTCGCGAGCAGCGCGCTCT 77
644 oProAlaProProValArgAlaThrThrProProProAlaProProAlaP 661
76 GCGCGGCTTCTCTGCGCTGGTGGTGGCG.....CGCGCGTGGCGCGCC 33
661 roProAlaProAsnSerMetAlaLeuProProAlaProProAspProPro 677

seq_name: sp_rodent:070495

seq_documentation_block:
ID 070495 PRELIMINARY; PRT; 897 AA.
AC 070495;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PLENTY-OF-PROLINES-101.
GN SRRM1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Vaysiere B.M., Camonis J.H.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF062655; AAC17422.1;
DR MGD; MGI:1858303; Srrm1.
DR InterPro; IPRO02483; PWI.
DR Pfam; PF01480; PWI; 1.
DR SMART; SM00311; PWI; 1.
SQ SEQUENCE 897 AA; 101166 MW; 62160BEB6772BD10 CRC64;
```

## alignment\_scores:

Quality: 192.00 Length: 563  
Ratio: 0.869 Gaps: 27  
Percent Similarity: 39.254 Percent Identity: 25.933

## alignment\_block:

US-09-462-480-1/rev x 070495

Align seg 1/1 to: 070495 from: 1 to: 897

```
1271 TCGCCGCGCATGACAACCTCTCAGAGTGGCTCAACGCTATAACACGAG 1222
1272 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
282 SerArgSerArgThrArgSerArgSerProSerHisThrArgProArgAr 298
1221 AAAGGGGAGACGACGAGGAGTGCAGAACTCGCCCGCATCCCGTGTTCGCT 1172
1222 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
298 gArgHisArgSerArgSerArgSerTyrSerProArgArg.....Argp 313
1171 ATTTACGCGAATCGGGTTCCTATGCGAATCCAGTAGAGTTGC 1122
1172 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
313 rSerProArgArgArgProSerProArgArgThrPro..... 326
1121 CTTCGGTGAAGCATTCCTCGACCGCTTCGCTGATCGCCGCGCAGG 1072
1122 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
327 ProArgArgMetProProProProArgHisArgSerArgSerProGl 343
1071 T..... 1067
1068 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
343 YArgArgArgArgSerSerAlaSerLeuSerGlySerSerSers 360
1066 CAGCGGTTGTTACAGTCGGTAGCGGTGGCGTCCCATTTTCGTGGACAC 1017
1067 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
360 erSerSerArgSerArgSerPro..... 368
1016 CCTGGTACGCTCGAAGCGGTACCGC.....CCAGCGCGCTCGGAGC 973
1017 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
369 ProLysLysProProLysArgThrSerSerProProArgLysThrArgAr 385
972 TTGTCAGGAGTCTCTCCCTCGCTCAAGGAGGAATGAATGACGTGAC 923
923 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
385 gLeuSerProSerAlaSerProProArgArgArgHisArgProSer..... 400
922 ATTTCCCTGGATTGGCTTGGCGGCGCTCGATAC.....CCGCGA 882
923 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
401 ..SerProAlaThrProProProLysThrArgHisSerProThrProGln 416
881 AATTCCTACTGCTCTCTCATGTTTTTGGCTCCGTTCT...TTTCGTAT 835
835 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
417 GlnSerAsnArgThrArgLysSerArgValSerValSerProGlyArgTh 433
834 TAGCGGTCAGAACCCATTTCCGAGGACA.....G 804
804 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
433 rSerGlyLysValThrLysHisLysGlyThrGluLysArgGluSerProS 450
803 CGCCTGCTGCTCCTCTGT..... 784
784 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
450 erProAlaProLysProArgLysValGluLeuSerGluSerGluGluAsp 466
784 ..... 784
467 LysGlySerLysMetAlaAlaAspSerValGlnGlnArgArgGlnTy 483
784 ..... 784
483 rArgArgGlnAsnGlnGlnSerSerSerAspSerGlySerSerSerThrS 500
783 .....CGGCC.....TCAGATTGGA 766
766 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
500 erGluAspGluArgProLysArgSerHisValLysAsnGlyGluValGly 516
765 CGCCGG..... 760
760 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
517 ArgArgArgArgLeuSerProSerArgSerAlaSerProSerProArgLy 533
760 ..... 760
533 sArgGlnLysGluThrSerProArgMetGlnMetGlyLysArgTrpGlnS 550
759 ..CCTGACGAATATTCGTGAGATCTCGTAGTTCCTGCTTCGCTTAT 712
712 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

seq\_name: sp\_human:Q9UQ36

seq\_documentation\_block:

ID Q9UQ36 PRELIMINARY; PRT; 1275 AA.

AC Q9UQ36;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

```
550 erProValThrLysSerSerArgArgArgArgSerProSerProPro 566
711 TGGCTGCTTCTTGAAGC.....GCACACCGCGCTCGCGCC 671
712 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
567 AlaArgArgArgArgSerSerProSerProAlaProProProProPr 583
670 GTCCCCCGCGCGCGCCACTGCGCCCTGCA.....ACGAACC 633
633 | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
583 oProProProArgArgArgSerProThrProProProArgArgT 600
632 TCCCGTCCACTCCACCTGGTGGATCTGGTTTTCAGTCCCGGAGATCC 583
633 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
600 hrProSerProProProArgArgSerProSerProArgArgTyrSer 616
582 .....GCTCGAAA 575
617 ProProLleGlnArgArgTyrSerProSerProProLysArgArgTh 633
574 TTACTGCTCTCTGCGGAGGTAGCGCATCGGTCTTCTCATCTCTGCCAT 525
525 : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
633 rAlaSerProProProProLysArgArgAlaSerProSerProPro. 649
524 GCTGGACTACTTCTCTCTTACCTTCTCGCCCAAAATGTTGGCAAGTCT 475
475 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
650 .....ProLysArgArgValSerHisSer 657
474 TCCGGCGCGGTGGCGGGAAGTCTGTGTCATTACGGGAGCTCACCACT 425
425 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
658 ProProProLysGlnArgSerProThrValThrLysArgArgSerPro.S 674
424 CGTCTCTTCTGCTCCAGTCGCTCGTCTTCTTTCACGCTCTCGCGCG 375
375 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
674 erLeuSerSerLysHisArgLysGlySerSerProGlyArgSer.ThrAr 690
374 AGCGGTGCCGCGGACGACACCGCGGTGGTGGAGCGCGGATTCGGA 325
325 | : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :
690 gLluAlaArgSer...ProGlnProAsn.....LysArgHisSerProS 704
324 ACCCTGGCCCATCGCTCCGCGAC.....CCACCGAGCGCGCGC 287
287 : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
704 erProArgProArgAlaProGlnThrSerSerProProProValArg 720
286 .....CACCGTCACCGACATCCGGCAACAGCGCGCGGATCACCGAG 243
243 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
721 GlyAlaSerAlaSerProGlnGlyArgGlnSerPro.....SerProSe 735
242 GGGCAACCGCGCTTTCGATCAGCTGAGACA..... 212
212 : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
735 rThrArgProIleArgArgValSerArgThrProGluProLysLysIleL 752
211 ..TCAGCGCGTGGGTCAACGACCCACCTCGCCAGGTAGCGACTCCG 164
164 : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
752 ysLysAlaAlaSerProSerProGlnSerValArgArgValSerSer 768
163 CGCGCAGCAGG.....CCGCGCGCGCTGGGGCTGATCCACACAGC 121
121 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
769 ArgSerValSerGlySerProGluProAla...AlaLysLysProProAl 784
120 CAGCGGATGTTTCGACAGCGGACTGGTCCGAGCAGGCGCCATCTCGCGG 71
71 | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
784 a.....ProSerProValGlnSerG 792
70 CTTCCTCGTCCGCTGGGTTCGCGCGCGGTGCGC 36
36 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
792 lnSerProSerThrAsnTrpSerProAlaValPro 803
```

DT 01-DEC-2001 (TREMBLrel.19, Last annotation update)  
DE RNA BINDING PROTEIN (FRAGMENT).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID:9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Ohtaki S., Umeki K., Sawada Y.;  
RT "Homo sapiens mRNA for RNA binding protein, partial cds.";  
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB016091; BAA83717.1; -;  
FT NON\_TER  
FT 1  
SQ SEQUENCE 1275 AA; 136869 MW; 45C2B2F85E98A6F6 CRC64;

alignment\_scores:  
Quality: 189.00 Length: 492  
Ratio: 0.875 Gaps: 21  
Percent Similarity: 43.902 Percent Identity: 26.829

alignment\_block:  
US-09-462-480-1 x Q9UQ36

Align seg 1/1 to: Q9UQ36 from: 1 to: 1275

5 AGCAGGTGACGTCGTGTTCAGCCAGGTGGCGGCACCGCGCGGCAAC 54  
|||||  
188 SerArgThrAlaArgGlySerArg...SerSerProGluProLysTh 203  
55 CCAGCGCAGGAGGAGCGCGCAGATGGCGCTGC...TCGGCACCATCC 101  
|||||  
203 rLysSerArgThrProProArgArgArgSerArgSerSerProGlu 220  
102 GCTGTCGAACCATCGCTGGCTGGTATCAGCCCGCCCGCGCGG 151  
|||||  
220 euThrArgLysAlaArg...LeuSerArgArgSerArg 231  
152 GCCTGCTGCGCGGAGTTCGTACTGCGCAGGTGGTGTGCTGACCCG 201  
|||||  
232 SerAlaSerSerProGlu... 238  
202 ACGCCGTGATGTCAGCTGATCGAAAGCCGTTGCCCTCGGTGAT 251  
|||||  
239 .....ThrArgSerArgThrProProArg... 246  
252 GCCGCGCTGCTCCG... 268  
247 ..HisArgSerSerProSerValSerSerProGluProAlaGluLys 262  
269 .....GATCGTCGGTGAAGGTGGCGCGCTCCCGTGGTCCGGAGCG 312  
|||||  
263 ArgSerSerArgArgArgArgSerAlaSerSerProArgThrLysTh 279  
313 ATGGCGCAGGTTCCGAATCGCGCGGTCCACACCGCCGCGTGTGTCG 362  
|||||  
279 rSerArgGlyArgSerProSerProLysProArgLysLeuLys 296  
363 GCCGCGCAGGTCGCGCAGGCGGTGAAGAGACGAGGAGGAGGAGG 412  
|||||  
296 erArgSerArgArgArgGluLysThrArgThrThrArgArg 312  
413 ACGAAGAGGAGCTGCTGCTCCCTTAATGACACAGAGCTCCCGGCC 462  
|||||  
312 pArgSerGlySerSerGlnSerThrSerArgArgArgGlnArgSer 329  
463 ACCGCGCGGAGAGCTGCCAATGCTGGCGAGGAGGAGGAGGAGAGA 512  
|||||  
329 erArgSerArgValThrArgArgArgArgGlyGlySerGlyTyHis 345  
513 AAGTAGTCCAGCATGGCAG... 532  
346 Arg.SerProAlaArgGlnGluSerSerArgThrSerSerArgArg 362

533 GATGAAGACCGATGCCCTACCTCGGCGCAGGAGGAGGTAATTCGAGC 582  
|||||  
362 rGlyArgSerArgThrProProThrSerArgLysArgSerArgSer 378  
|||||  
583 GGATCTCGGCGACTGAAGA... 603  
|||||  
379 ThrSerProAlaProTriLysArgSerArgSerArgAlaSerProAla 395  
604 .....CCCATGTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 637  
|||||  
395 rHisArgArgSerArgSerArgThrProLeuLysSerArgArgSer 412  
638 GTTCGAGGCGCAGTGGCGCGCGGCGGAGGAGGAGGAGGAGGAGG 684  
|||||  
412 rGSerArgThrSerProValSerArgArgArgSerArgSerArgThr 428  
685 .....CGTGGTGGCTTCCAAAGCAGCAGCAGCAGCAGCAGCAG 725  
|||||  
429 ValThrArgArgSerArgSerArgAlaSerProValSerArgArg 445  
726 GAATCGCAGGAGATCTCGACGAATATTCGTGAGGCGCGGCGTCCA 775  
|||||  
445 gSerArgSerArgThrProProVal...ThrArgArgSerArgSer 461  
776 GAGGCGCGCAGC...AGGAGCAGCAGCAGGCGCTGCTCT... 810  
|||||  
461 rGThrProThrArgArgArgSerArgSerArgThrProProValThr 477  
811 ...CGCAATGGGCTTCTGA...CCCGCTAATAGCAAGAAAGACGG 850  
|||||  
478 ArgArgArgSerArgSerArgThrProProValThrArgArgSer 494  
851 AGCAAAACA...TGACAGCAGCAGCAGTGAATTCGC 885  
|||||  
494 gSerArgThrSerProThrArgArgSerArgSerArgSerArgThr 511  
886 GGGTATCGAGCGCGCGCAGCGCA...TCCAGGGAA 920  
|||||  
511 roValThrArgArgArgSerArgSerArgThrSerProValThrArg 527  
921 ATGTCAGCTCATTCATTCCTCTCTGACGAGGAGGAGGAGGAGG 970  
|||||  
528 ArgSerArgSerArgThrSerProValThrArgArgSer... 541  
971 AAGCTCGCAGCGCTGGCGGTAGCGGTTCGAGCGGTACCGAGGTGT 1020  
|||||  
542 .ArgSerArgThrProProAlaLysArgArgSerArgSerArg... 556  
1021 CCAGCAAAATGGGAGCGCGCTACCGAGCTGAACACCGCGCTGC... 1067  
|||||  
557 .....ThrProLeuLeuProArgLysArgSerArgSerArg 568  
1068 .....AGAACCTG 1075  
569 SerProLeuAlaLysArgArgSerArgSerArgThrProArgThrAl 585  
1076 GCGCGGA...CGATCAGGAGCGCGTCCAGCAATGGCTCCGACCGAGG 1122  
|||||  
585 aArgGlyLysArgSerLeuThrArgSerProProAlaLysArgArg 602  
1123 CAAGCTCAGCTGGATGTCGCATAGGCGCAGCGGAGTTCGCTAGATA 1172  
|||||  
602 erAlaSerGlySerSerSerArgSerArgSerAlaThrProProAla 618  
1173 GCGAAACACCGGATCGGCGAGTTCGACCTCCGCTCGGCTCGGCTTC 1222  
|||||  
619 ThrArgAsnHisSerGlySerArgThrProProValAlaLeuAsnSer 635  
1223 TCGTGTGTTATACGTTTCGAGCGC 1244  
635 rArgMetSerCysPheSerArg 642

seq\_name: sp\_human:Q9Y2J6

seq\_documentation\_block:

ID Q9Y2J6 PRELIMINARY; PRT; 772 AA.  
AC Q9Y2J6;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE KIAA0992 PROTEIN (FRAGMENT).  
GN KIAA0992.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=BRAIN;  
RX MEDLINE=99246063; PubMed=10231032;  
RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Hirose M.,  
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;  
RT "Prediction of the coding sequences of unidentified human genes. XIII.  
RT The complete sequences of 100 new cDNA clones from brain which code  
RT for large proteins in vitro";  
RL DNA Res. 6:63-70(1999).  
DR EMBL; AB023209; BAA76836.1; -.  
DR HSSP; P56276; ITLK.  
DR InterPro; IPR003598; Ig\_C2.  
DR InterPro; IPR003006; Ig\_MHC.  
DR InterPro; IPR002965; P-rich\_extensn.  
DR Pfam; PF00047; 19; 3.  
DR PRINTS; PR01217; PRICHEXTENS.  
DR SMART; SM00408; IGC2; 3.  
KW Immunoglobulin domain.  
FT NON\_TER 1  
SQ SEQUENCE 772 AA; 83643 MW; EE5EE2BC79C7492B CRC64;

alignment\_scores:

Quality:	185.00	Length:	312
Ratio:	1.381	Gaps:	15
Percent Similarity:	42.949	Percent Identity:	28.526

alignment\_block:

US-09-462-480-1/rev x Q9Y2J6 ..

Align seg 1/1 to: Q9Y2J6 from: 1 to: 772

```

822 AGCCATTTCGAGG.....ACAGCGCTGCTGCTGCTCC 788
||||| :|||
3 SerProGlyGlyArgGlyIleLysProAspThrCysProAlaProGlyPr 19
787 TCGTCGGCCC.....:.....TCGAGTA 771
||||| |||
19 oArgSerProLeuGlnLeuProLeuAlaProAspAlaGluSerGlySers 36
770 TTGGACGCGCGCTGAC.....:..... 754
||||| |||
36 erGlyArgArgProGlyCyluProArgaspProLeuLysLeuGlnGlnLeu 52
753 .....GAATATTCGTGAGATCTCGTCGAGTTCTCTGC 722
||||| :|||
53 GlnAsnGlnIleArgLeuGluGlnGluAlaGlyAlaArgGlnProProPr 69
721 TTCGTCTATTGGCTGCTTCTTGAAGCGCACCGCGCGCTGGCGGC 672
::: ||||| |||
69 oAlaPro.....ArgSerAlaProProSerProProPheP 81
671 CGTCCCGCGCGCGCGCCACTGGCCCTGCAACGAACTGCGCTGACT 622
|| ||||| |||
81 roProProAlaPheProGluLeuAlaAlaCysThrProProAlaSer 97
621 CCACCTGCTGCTGGGTTTTCAGTTCGCGGAGATCCGCTCGA..... 577
||| ||||| |||

```

```

98 ProGluProMetSerAlaLeuAlaSerArgSerAlaProAlaMetGlnSe 114
576 .....AATTACTGCTCTCCGCCGAGGGTAGCGCATCGGT 540
||||| ||| :|
114 rSerGlySerPheAsnTyAlaArgProLysGln..... 125
539 CTTTCATCTCTGCCATGCTGGACTACTTTCTCTCTTTACCTTTCCTCGCAA 490
::: ||| :|||
126 .....PheileAlaAlaGln 130
489 AATGTTGGCACTCTTCGGGCC..GGTGGCGGGGAAGTCTGTCTCATTT 441
||||| :||| ||| ||||| |||
131 AsnGlyProAlaSerGlyHisGlyThrProAlaSerProSerSe 147
440 ACGGAGCTCACAGTCTCTCTCTCGTCCAGTCTCTCTCTCTCTCTCTTC 391
::||| ||| ||||| :||| :||| :||| :|||
147 rSerSerLeuProSer....ProMetSerProThr....ProArgGlnPheG 162
390 TTCACGCTCTCT.....GCGGAGCGGTGCCGCGCGACAGAC 353
::||| ||| :||| :||| :||| :||| :|||
162 LyArgAlaProValProPheAlaGlnProPheGlyAlaGluProGlu 178
352 CCGGCTGCTGAGCGCGCGGTTCGGAACCTGCGCCCATCTCTCTCCGGA 303
||| ||| :||| :||| :||| :||| :|||
179 AlaProTrpGlySer.....SerSerProSerProProProPr 192
302 CCCA.....CGGAGCGCGCCACCGCTCACCGACGATCCGGC 265
||| ||| ||| ||| ||| |||
192 oProValPheSerProThrAlaAlaPheProValProAspValPheP 209
264 AACAGCGCGCGCATCACCGAGGGGCAACCGGCTTTTCGATCAGCTGAG 215
||||| |||
209 roLeuProProProPro..... 215
214 ACATCAGCGCGGTGCGGTCAACGACCCACCTGCGCCAGTAGCGACTCC 165
||| :||| :||| :||| :||| :|||
216 .....ProLeu..ProSerProGlyGlnAlaSer 224
164 GCGCGCAGCAGCGCGCGCGCTGGG...CCTGATCCACACGACGAG 118
||| ||||| :||| :||| :||| :||| :|||
225 HisCysSerSerProAlaThrArgPheGlyHisGlyGlnThrProAlaAl 241
117 CGGATGTTGACAGCGGACTGGTCCGAGCAGGCGCCATCTCTCGCGGCTT 68
: ||| ||||| :||| :||| :||| :||| :|||
241 a.....PheLeuSerAlaLeuLeuProSerGlnProProProAlaAlav 256
67 CCGTCTGCGGTGGTTCGCGCGCGCGGTGGCGGCC 33
::: ||||| ||| |||
256 alAsnAlaLeuGlyLeuProLysGlyValThrPro 267

```

